



Universidade de Aveiro Departamento de Química
2014

**Sílvia de Oliveira Diaz Doenças da gravidez e do bebé estudadas por
metabolómica de urina**

**Pregnancy and newborns disorders followed by
urine metabolomics**



Universidade de Aveiro Departamento de Química
2014

Sílvia de Oliveira Diaz Doenças da gravidez e do bebé estudadas por metabolómica de urina

Pregnancy and newborns disorders followed by urine metabolomics

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Química, realizada sob a orientação científica da Doutora Ana Maria Pissarra Coelho Gil, Professora Associada com Agregação do Departamento de Química da Universidade de Aveiro e do Doutor António de Sousa Barros, Investigador Auxiliar da Unidade de Investigação de Química Orgânica, Produtos Naturais e Agro-alimentares (QOPNA) da Universidade de Aveiro

Apoio financeiro da Fundação para a Ciência e Tecnologia (FCT) - bolsa de Investigação FCT SFRH/BD/64159/2009 financiada pelo Programa Operacional Potencial Humano (POPH) e projecto PTDC/QUI/66523/2006 financiado pelo Fundo Europeu de Desenvolvimento Regional (FEDER) e pelo Programa Operacional Factores de Competitividade (COMPETE); da Universidade de Aveiro - Centro de Investigação em Materiais Cerâmicos e Compósitos (CICECO); da empresa Bruker BioSpin GmbH; Agradecemos à Rede Nacional de RMN (RNRMN), suportada com fundos da FCT.



Dedicated to my best friends, my parents.

o júri
presidente

Professor Doutor Helmuth Robert Malonek
Professor Catedrático do Departamento de Matemática da Universidade de Aveiro

Doutora Laura Vilarinho
Investigadora Auxiliar Responsável da Unidade de Rastreio Neonatal, Metabolismo e Genética do Departamento de Genética Humana Instituto de Saúde Dr. Ricardo Jorge, INSA

Professor Doutor Carlos Frederico de Gusmão Campos Geraldês
Professor Catedrático do Departamento Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

Professor Doutor Hugh Douglas Burrows
Professor Catedrático do Departamento de Química da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

Doutora Isabel Marques Carreira
Professora Auxiliar com Agregação da Faculdade de Medicina da Universidade de Coimbra (FMUC), Directora do Laboratório de Citogenética e Genómica - FMUC, Coordenadora do Centro de Investigação Meio Ambiente, Genética e Oncobiologia (CIMAGO) da FMUC e membro do Centro de Ciências Forenses (CENCIFOR) do Instituto de Medicina Legal

Doutor Brian James Goodfellow
Professor Auxiliar do Departamento de Química, Laboratório Associado CICECO da Universidade de Aveiro

Doutora Ana Maria Pissarra Coelho Gil
Professora Associada com Agregação do Departamento de Química, laboratório Associado CICECO da Universidade de Aveiro

Doutor António de Sousa Barros
Investigador Auxiliar da Unidade de Investigação de Química Orgânica, Produtos Naturais e Agro-alimentares (QOPNA) da Universidade de Aveiro

Acknowledgements

I would like to express my sincere gratitude to my supervisors Dr. Ana M. Gil and Dr. António S. Barros for their guidance and all the knowledge they passed on to me. I am very thankful to them for introducing me to the field of metabolomics and for giving me the opportunity to pursue this exciting work that I have become so fond of.

I want to acknowledge all the team and entities involved in this research project without whom this work would not have been possible. I particularly want to thank Dr. Maria do Céu Almeida, Dr. Eulália Galhano, Dr. Fátima Negrão and Nurse Cristina Pita (from Maternidade Bissaya Barreto, Coimbra) and Dr. Isabel Marques Carreira (from Laboratório de Citogenética e Genómica, Faculdade de Medicina, Universidade de Coimbra) for their continuous and relentless dedication to this research. I also express my gratitude to Dr. Iola F. Duarte and Dr. Brian J. Goodfellow (from University of Aveiro) for their help with the NMR experiments and biochemical inputs. I thank the students that work with me, Elisabete Morais and Elisabete Aguiar, for their contribution. Lastly, I want to acknowledge my project colleagues, Gonçalo Graça and Joana Pinto, for their invaluable help and fellowship over these years.

I would also like to acknowledge the Biomolecular Medicine group of Imperial College of London for providing access to the UPLC-MS experiments. I express my gratitude for the opportunity to work with this notable group, especially to Prof. John C. Lindon, Prof. Elaine Holmes, Dr. Maria Gomez-Romero, Dr. Matthew R. Lewis, Dr. Kirill Veselkov, and Dr. Paul H. Benton. I also thank Dr. Manfred Spraul (from Bruker BioSpin, Germany) for providing access to spectral databases and performing some NMR experiments.

I acknowledge the Portuguese National NMR Network (RNRMN), supported with funds from the Foundation for Science and Technology (FCT), and financial support from the European Regional Development Fund through the Competitive Factors Thematic Operational Program, FCT, CICECO and University of Aveiro, for providing the means necessary to successfully conclude this work.

I feel very lucky to have worked with such an extraordinary group, that I now call my friends. I especially want to mention Cláudia Rocha, Gonçalo Graça, Inês Iamego, Joana Carrola, Joana Marques, Joana Pinto, João Rodrigues, Sérgio Vilela and Susana Aveiro. I could never thank them enough for their friendship, motivation and the countless laughs and moments we have shared. These years were undoubtedly better because of them.

I am very fortunate to have so many wonderful people in my life. I am specially grateful to have shared this journey with my dearest Miguel and for his unwavering companionship, support and patience.

I am also thankful to my siblings for, probably without knowing, having always inspired me to go further and do better.

The best part comes always at the end. I proudly want to share this accomplishment with my parents without whom I simply could not have achieved this milestone. Thanking them would just not be enough. This is also yours.

Palavras-chave

Urina materna, urina do bebê, saúde pré-natal, metabolómica, espectroscopia de ressonância magnética nuclear (RMN), cromatografia líquida de ultra eficiência acoplada a espectrometria de massa (UPLC-MS), análise multivariada, selecção de variáveis, malformações fetais, trissomia 21, parto pretermo, diabetes mellitus gestacional, pré-eclampsia, restrição do crescimento intrauterino.

Resumo

O Capítulo 1 descreve o enquadramento deste trabalho identificando as doenças pré-natais relevantes e os métodos de diagnóstico actualmente disponíveis. É depois apresentada a metodologia seguida, assim como uma breve introdução dos princípios dos métodos analíticos e estatísticos aplicados. O capítulo é concluído com uma descrição do estado da arte na área de metabolómica em investigação pré-natal, identificando o mérito desta inovadora estratégia para a identificação de marcadores robustos de doenças pré-natais. A relevância deste trabalho torna-se clara através do escasso uso de urina materna e do recém-nascido em trabalhos anteriores. O Capítulo 2 descreve os procedimentos experimentais utilizados neste trabalho, incluindo condições de amostragem, recolha e preparação das amostras, protocolos de aquisição e de tratamento dos dados.

A caracterização da composição da urina materna, através de espectroscopia de Ressonância Magnética Nuclear (RMN) de próton é apresentada no Capítulo 3. Define-se o perfil metabólico urinário característico para cada trimestre de gravidez, tendo sido encontrado um conjunto de 21 metabolitos descritivo das alterações metabólicas ocorridas ao longo da gravidez. 8 metabolitos foram encontrados a variar com a gravidez, pela primeira vez, tendo sido confirmadas variações metabólicas conhecidas. É ainda estudado o efeito do não-jejum (usado neste trabalho) como possível factor de confusão.

O Capítulo 4 apresenta o estudo metabolómico de urina materna do 2º trimestre para o diagnóstico de doenças fetais e previsão de complicações mais tarde desenvolvidas. Este estudo compreende a aplicação de um método de selecção de variáveis desenvolvido no âmbito desta tese. Observou-se que as malformações fetais (e, especificamente, do sistema nervoso central, SNC) e as cromossomopatias (e, especificamente, a trissomia 21, T21) são acompanhadas por alterações nos metabolismos energético, dos aminoácidos, lípidos e nucleótidos, enquanto que as cromossomopatias mostraram ser acompanhadas por uma desregulação adicional dos metabolismos dos açúcares, ciclo da ureia e/ou biossíntese da creatinina. A validação dos modelos multivariados revelou taxas de classificação (CR) de 84% para malformações (87%, SNC) e 85% para CD (94%, T21). Para o parto pré-termo, pré-eclampsia (PE) e restrição de crescimento intrauterino (RCIU) observaram-se perfis que podem ajudar à previsão precoce, com CR 84% para pretermo (11-20 semanas de gestação, g.w. pré-diagnóstico), 94% para PE (18-24 g.w. pré-diagnóstico) e 94% para RCIU (2-22 g.w. pré-diagnóstico). Este capítulo inclui resultados obtidos por cromatografia líquida de ultra eficiência acoplada a espectrometria de massa (UPLC-MS) para pré-pretermo e correlação com os dados de RMN. Um possível composto marcador foi detectado mas a sua identificação não foi possível.

O Capítulo 5 descreve o estudo metabolómico por RMN da diabetes mellitus gestacional (DMG), estabelecendo-se um perfil metabólico potencialmente preditivo da doença (CR 83%, 2-21 g.w. pré-diagnóstico). Verificou-se ainda que o espectro de RMN contém informação sobre o fenótipo individual, capaz de prever a necessidade futura de tratamento com insulina (CR 94%).

No Capítulo 6 demonstra-se o impacto do tipo de parto (CR 88%) e género do bebé (CR 76%) no perfil da urina do recém-nascido. Verificou-se ainda que a prematuridade, depressão respiratória, crescimento grande para a idade gestacional e malformações induzem perturbações metabólicas relevantes (CR 82-92%), assim como algumas doenças maternas como a DMG (CR 82%) e doenças psiquiátricas (91% CR).

Finalmente, no Capítulo 7 apresentam-se as principais conclusões deste trabalho, enfatizando o potencial da metabolómica de urina materna e do bebé para o acompanhamento da gravidez e previsão de doenças, visando o desenvolvimento de novos métodos de diagnóstico precoce e não-invasivo.

Keywords

Maternal urine, newborn urine, prenatal health, metabolomics, nuclear magnetic resonance (NMR), ultraperformance liquid chromatography-mass spectrometry (UPLC-MS), multivariate analysis (MVA), variable selection, fetal malformations (FM), trisomy 21 (T21), preterm delivery (PTD), gestational diabetes mellitus (GDM), preeclampsia (PE), intrauterine growth restriction (IUGR).

Abstract

Chapter 1 introduces the scope of the work by identifying the clinically relevant prenatal disorders and presently available diagnostic methods. The methodology followed in this work is presented, along with a brief account of the principles of the analytical and statistical tools employed. A thorough description of the state of the art of metabolomics in prenatal research concludes the chapter, highlighting the merit of this novel strategy to identify robust disease biomarkers. The scarce use of maternal and newborn urine in previous reports enlightens the relevance of this work. Chapter 2 presents a description of all the experimental details involved in the work performed, comprising sampling, sample collection and preparation issues, data acquisition protocols and data analysis procedures.

The proton Nuclear Magnetic Resonance (NMR) characterization of maternal urine composition in healthy pregnancies is presented in Chapter 3. The urinary metabolic profile characteristic of each pregnancy trimester was defined and a 21-metabolite signature found descriptive of the metabolic adaptations occurring throughout pregnancy. 8 metabolites were found, for the first time to our knowledge, to vary in connection to pregnancy, while known metabolic effects were confirmed. This chapter includes a study of the effects of non-fasting (used in this work) as a possible confounder.

Chapter 4 describes the metabolomic study of 2nd trimester maternal urine for the diagnosis of fetal disorders and prediction of later-developing complications. This was achieved by applying a novel variable selection method developed in the context of this work. It was found that fetal malformations (FM) (and, specifically those of the central nervous system, CNS) and chromosomal disorders (CD) (and, specifically, trisomy 21, T21) are accompanied by changes in energy, amino acids, lipids and nucleotides metabolic pathways, with CD causing a further deregulation in sugars metabolism, urea cycle and/or creatinine biosynthesis. Multivariate analysis models' validation revealed classification rates (CR) of 84% for FM (87%, CNS) and 85% for CD (94%, T21). For later-diagnosed preterm delivery (PTD), preeclampsia (PE) and intrauterine growth restriction (IUGR), it is found that urinary NMR profiles have early predictive value, with CRs ranging from 84% for PTD (11-20 gestational weeks, g.w., prior to diagnosis), 94% for PE (18-24 g.w. *pre*-diagnosis) and 94% for IUGR (2-22 g.w. *pre*-diagnosis). This chapter includes results obtained for an ultraperformance liquid chromatography-mass spectrometry (UPLC-MS) study of *pre*-PTD samples and correlation with NMR data. One possible marker was detected, although its identification was not possible.

Chapter 5 relates to the NMR metabolomic study of gestational diabetes mellitus (GDM), establishing a potentially predictive urinary metabolic profile for GDM, 2-21 g.w. prior to diagnosis (CR 83%). Furthermore, the NMR spectrum was shown to carry information on individual phenotypes, able to predict *future* insulin treatment requirement (CR 94%).

Chapter 6 describes results that demonstrate the impact of delivery mode (CR 88%) and gender (CR 76%) on newborn urinary profile. It was also found that newborn prematurity, respiratory depression, large for gestational age growth and malformations induce relevant metabolic perturbations (CR 82-92%), as well as maternal conditions, namely GDM (CR 82%) and maternal psychiatric disorders (CR 91%).

Finally, the main conclusions of this thesis are presented in Chapter 7, highlighting the value of maternal or newborn urine metabolomics for pregnancy monitoring and disease prediction, towards the development of new early and non-invasive diagnostic methods.

List of publications including the work presented in this thesis

Diaz SO, Pinto J, Graça G, Duarte IF, Barros AS, Galhano E, Pita C, Almeida MC, Goodfellow BJ, Carreira IM, Gil AM. Metabolic biomarkers of prenatal disorders: an exploratory NMR metabonomics study of 2nd trimester maternal urine and blood plasma. *Journal of Proteome Research*, **2011**, 10(8), 3732–3742.

Graça G, Diaz SO, Pinto J, Barros AS, Duarte IF, Goodfellow BJ, Galhano E, Pita C, Almeida MC, Carreira IM and Gil AM. Can biofluids metabolic profiling help to improve healthcare during pregnancy?, *Spectroscopy: An International Journal*, **2012**, 27(5-6): 515-523.

Diaz SO, Barros AS, Goodfellow BJ, Duarte IF, Carreira IM, Galhano E, Pita C, Almeida MC and Gil AM. Following healthy pregnancy by nuclear magnetic resonance (NMR) metabolic profiling of human urine. *Journal of Proteome Research*, **2013**, 12(2), 969-979.

Diaz SO, Barros AS, Goodfellow BJ, Duarte IF, Galhano E, Pita C, Almeida MC, Carreira IM and Gil AM. Second trimester maternal urine for the diagnosis of trisomy 21 and prediction of poor pregnancy outcomes. *Journal of Proteome Research*, **2013**, 12(6), 2946-2957.

Manuscripts in preparation

Diaz SO, Aguiar E, Barros A, Goodfellow BJ, Duarte IF, Galhano E, Pita C, Almeida MC, Carreira IM, Barros AS and Gil AM, “Urine profiling can predict onset of gestational diabetes and treatment response” *in preparation*.

Diaz SO, Morais E, Goodfellow BJ, Duarte IF, Negrão F, Pita C, Almeida MC, Carreira IM, Barros AS and Gil AM, “Urine metabolomics for newborn health assessment: I. Newborns diseases”, *in preparation*.

Diaz SO, Morais E, Goodfellow BJ, Duarte IF, Negrão F, Pita C, Almeida MC, Carreira IM, Barros AS and Gil AM, “Urine metabolomics for newborn health assessment: II. Maternal diseases”, *in preparation*.

Index

1. Introduction	29
1.1 Healthy pregnancy and fetal development	29
1.2 Prenatal disorders and diagnostic procedures: availability and needs	32
1.2.1 Fetal Malformations	33
1.2.2 Chromosomal Disorders	36
1.2.3 Intrauterine Growth Restriction	43
1.2.4 Preeclampsia	44
1.2.5 Preterm Delivery	45
1.2.6 Gestational Diabetes Mellitus	46
1.2.7 Premature Rupture of the Membranes	48
1.3 Metabolomics: envisaging new clinical strategies in prenatal health	49
1.4 Principles of Nuclear Magnetic Resonance (NMR) spectroscopy and Mass Spectrometry (MS)	53
1.4.1 NMR Spectroscopy	53
1.4.2 Mass Spectrometry	61
1.5 Statistical tools in metabolomics	63
1.5.1 Preprocessing: bucketing, alignment, normalization, scaling and transformation	64
1.5.2 Multivariate analysis methods	67
1.5.3 Univariate statistical methods	72
1.5.4 Variable selection methods	74
1.5.5 Validation methods	75
1.6 State of the art of metabolomics in prenatal research	79
1.7 Aims and scope of this thesis	88
2. Experimental section	89
2.1 Sampling	89
2.1.1 Sample and metadata collection	89
2.1.2 Definition of sample groups	91

2.2 NMR spectroscopy	97
2.2.1 Sample preparation for NMR analysis	97
2.2.2 Acquisition and processing of NMR data	97
2.2.3 Preprocessing and multivariate analysis of NMR data	99
2.3 Ultraperformance liquid chromatography- mass spectrometry (UPLC-MS)	102
2.3.1 Sample preparation for UPLC-MS analysis	102
2.3.2 Acquisition and processing of UPLC-MS data	102
2.3.3 Preprocessing and multivariate analysis of UPLC-MS data	104
2.4 Statistical correlation analysis	105
<i>3. Defining the global urinary profile of pregnancy and the metabolic adaptations throughout gestation - a multipoint collection study</i>	<i>107</i>
3.1 Typical ^1H NMR spectrum of maternal urine and spectral assignments	107
3.2 Impact of non-fasting conditions on urine metabolomic studies	113
3.2.1 Effect of non-fasting on urine composition	114
3.2.2 Proposed metabolic interpretation of the urinary changes under non-fasting	115
3.3 Following healthy pregnancy metabolism by NMR metabolic profiling of maternal urine	116
3.3.1 Effect of healthy pregnancy on urine metabolite profile	116
3.3.2 Proposed metabolic interpretation of the urinary changes found across pregnancy	126
3.3.3 Summary of changes across pregnancy	131
<i>4. Second trimester maternal urine for the diagnosis of fetal disorders and prediction of poor pregnancy outcomes</i>	<i>133</i>
4.1 Impact of diagnosed fetal disorders on maternal urine composition	133
4.1.1 Fetal malformation (FM) cases	133
4.1.1.1 Development and validation of a variable selection methodology	134
4.1.1.2 Unveiling urinary metabolic signatures of general FM and central nervous system (CNS) malformations	138
4.1.1.3 Preliminary external validation	142

4.1.1.4 Proposed metabolic interpretation of urinary changes in general FM and CNS malformations	145
4.1.2 Chromosomal disorder (CD) cases	147
4.1.2.1 Unveiling urinary metabolic signatures of general CD and trisomy 21 (T21)	147
4.1.2.2 Proposed metabolic interpretation of the urinary signature of general CD and T21	151
4.2 Prediction of poor pregnancy outcomes through second trimester maternal urine metabolomics	155
4.2.1 Women later developing preterm delivery (PTD)	155
4.2.1.1 Unveiling a <i>pre</i> -diagnosis urinary metabolic signature of <i>pre</i> -PTD by NMR	155
4.2.1.2 Preliminary external validation of NMR models	158
4.2.1.3 Unveiling a <i>pre</i> -diagnosis urinary metabolic signature of PTD by UPLC-MS	159
4.2.1.4 Statistical correlation of NMR and MS data	164
4.2.1.5 Proposed metabolic interpretation of <i>pre</i> -diagnosis urinary changes in PTD	166
4.2.2 Women later developing preeclampsia (PE)	168
4.2.2.1 Unveiling a <i>pre</i> -diagnosis urinary signature of PE	168
4.2.2.2 Proposed metabolic interpretation of <i>pre</i> -diagnosis urinary changes in PE	171
4.2.3 Women carrying fetuses later diagnosed with intrauterine growth restriction (IUGR)	172
4.2.3.1 Unveiling a <i>pre</i> -diagnosis urinary metabolic signature of IUGR	172
4.2.3.2 Proposed metabolic interpretation of <i>pre</i> -diagnosis urinary changes in IUGR	175
4.2.4 Women later developing premature rupture of membranes (PROM)	177
4.3 Summary of the impact of fetal and maternal disorders	178

5. Gestational Diabetes Mellitus: a deeper insight into maternal metabolism disturbances	181
5.1 Maternal urine composition at the time of GDM diagnosis	181
5.1.1 Unveiling an urinary metabolic signature of GDM	181
5.1.2 Proposed metabolic interpretation of urinary changes in GDM	186
5.2 Women later developing GDM	188
5.2.1 Unveiling a <i>pre</i> -diagnosis urinary metabolic signature of GDM	188
5.2.2 Proposed metabolic interpretation of the <i>pre</i> -diagnosis changes in GDM	191
5.3 Impact of GDM treatment on urine metabolic composition	193
5.3.1 Unveiling urinary metabolic signatures of GDM treatment responses	193
5.3.2 Proposed metabolic interpretation of the effect of GDM treatments	198
5.4 Prediction of GDM treatment response: a possible pharmacometabolomic application	202
5.4.1 Unveiling a <i>pre-treatment</i> metabolic phenotype predictive of GDM treatment requirements	202
5.4.2 Proposed metabolic interpretation of <i>pre</i> -treatment urinary profiles	206
5.5 Summary of the impact of GDM in maternal urine	207
6. Newborn health status seen through urine metabolomics	211
6.1 Typical NMR urine spectra of newborn urine and assignments	211
6.2 Impact of delivery mode and gender in the urine of healthy newborns	215
6.2.1 Effects of delivery mode on urine composition: results and proposed metabolic interpretation	216
6.2.2 Effect of gender on urine composition: results and proposed metabolic interpretation	221
6.3 Impact of newborns disorders on newborn urine composition	227
6.3.1 Prematurity	227
6.3.1.1 Unveiling an urinary metabolic signature of premature newborns	228
6.3.1.2 Proposed metabolic interpretation of changes related with prematurity	231
6.3.2 Respiratory depression episode	237
6.3.2.1 Unveiling an urinary metabolic signature of respiratory depression	238

6.3.2.2 Proposed metabolic interpretation of changes related with respiratory depression	242
6.3.3 Large for Gestational Age (LGA)	243
6.3.3.1 Unveiling an urinary metabolic signature of LGA	244
6.3.3.2 Proposed metabolic interpretation of changes related with LGA	247
6.3.4 Malformations	249
6.3.4.1 Unveiling an urinary metabolic signature of malformations	249
6.3.4.2 Proposed metabolic interpretation of changes related with malformations	253
6.4 Impact of maternal disorders on newborn urine composition	254
6.4.1 Maternal Gestational Diabetic Mellitus (GDM)	254
6.4.1.1 Unveiling an urinary metabolic signature of GDM	255
6.4.1.2 Proposed metabolic interpretation of changes related with GDM	257
6.4.2 Maternal psychiatric disorders	259
6.4.2.1 Unveiling an urinary metabolic signature of maternal psychiatric disorders	260
6.4.2.2 Proposed metabolic interpretation of the changes related with maternal psychiatric disorders	262
6.5 Summary of newborns and maternal disorders in newborn urine	264
7. Conclusions and future perspectives	267
8. Bibliography	273
<i>Annex I: Main characteristics of prenatal and newborn health metabolomic studies</i>	293
<i>Annex II: Informed consents used in this study</i>	303
<i>Annex III: Multivariate model parameters and Monte Carlo Cross Validation (MCCV) results</i>	307
<i>Annex IV: Reproducibility of UPLC-MS experiments and model quality parameters</i>	313

List of abbreviations and symbols

1D	One dimensional
1 st T	1 st trimester
2D	Two dimensional
2-HIBA	2-hydroxyisobutyrate
2-KG	2-ketoglutarate
2 nd T	2 nd trimester
2-Py	N-methyl-2-pyridone-5-carboxamide
3-HBA	3-hydroxybutyrate
3-HIVA	3-hydroxyisovalerate
3 rd T	3 rd trimester
4-DEA	4-deoxyerythronic acid
4-DTA	4-deoxythreonic acid
4-HPA	4-hydroxyphenylacetate
4-Py	N-methyl-4-pyridone-3-carboxamide
ADA	American diabetes association
AF	Amniotic fluid
AFP	Alpha fetoprotein
AGA	Adequate for gestational age
ANN	Artificial Neural Networks
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
AUC	Area under the curve
BALF	Bronchoalveolar lavage fluid
BCAA	Branched chain amino acids
b _{cvSE}	b coefficient's standard error obtained by cross validation
BHMT	Betaine-homocysteine methyltransferase
b _i	b coefficient of variable i
BMI	Body mass index
br	Broad
BSLR	Backward stepwise logistic regression
CD	Chromosomal disorders
CE	Capillary electrophoresis
CHC	Centro Hospitalar de Coimbra
CICECO	Centro de Investigação em Materiais Cerâmicos e Compósitos
CMM	Correlation matrix map
CMV	Cytomegalovirus
CNS	Central nervous system
COSY	Correlation spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill
CR	Classification rate
CS	Cesarean section

CSF	Cerebrospinal fluid
CV	Cross validation
CVA	Canonical variates analysis
CVS	Chorionic villus sampling
d	Doublet
dd	Doublet of doublets
DMA	Dimethylamine
DMG	Dimethylglycine
DNA	Deoxyribonucleic acid
DS	Down syndrome
Efp	Exponential multiplication, fourier transform and phase correction
ERETIC	Electronic Reference To access In vivo Concentrations
ESI	Electrospray ionization
ESI-	Electrospray ionization in negative mode
ESI+	Electrospray ionization in positive mode
F	Female
F1	1 st dimension
F2	2 nd dimension
FAS	Fetal alcohol syndrome
FID	Free induction decay
FISH	Fluorescent insitu hybridization
FM	Fetal malformations
FMUC	Faculty of Medicine, University of Coimbra
FN	False negative
FP	False positive
FPG	Fasting plasma glucose
FPR	False positive rate
FT	Fourier transform
g.w.	Gestational weeks
GA	Gestational age
GAA	Guanidoacetate
GABA	γ -aminobutyrate
GAN	Group aggregating normalization
GC	Gas chromatography
GDM	Gestational diabetes Mellitus
GP	Genetic programming
GSH	Glutathione
HCA	Hierarchical cluster analysis
Hcy	Homocysteine
HELLP	Hemolysis, Elevated Liver enzymes and Low Platelets
HETCOR	Heteronuclear correlation
HET-STOCSY	Heteronuclear statistical total correlation spectroscopy

HIE	Hypoxic ischemic encephalopathy
HILIC	Hydrophilic interaction chromatography
HMBC	Heteronuclear Multiple Bond Correlation
HMDB	Human metabolome database
HMLR	Hierarchical multiple linear regression
HN	Histogram normalization
HPLC	High performance liquid chromatography
HSQC	Heteronuclear Single Quantum Coherence
HSS	High strength silica
IEM	Inborn errors of metabolism
iGDM	Infant from a GDM mother
INEPT	Insensitive nuclei enhanced by polarization transfer
iPLS	Interval Partial least squares
IR	Infrared spectroscopy
IS	Indoxyl sulphate
IUFD	Intrauterine fetal demise
IUGR	Intrauterine growth restriction
IVC	Interventricular communication
<i>J</i> -res	<i>J</i> -resolved
<i>k</i> NN	<i>k</i> -Nearest Neighbor
Lb	Line broadening
LBW	Low birthweight
LC	Liquid chromatography
LC-HRMS	Liquid chromatography high-resolution mass spectrometry
LDA	Linear discriminant analysis
LGA	Large for gestational age
LOESS	Locally weighted scatter plot smoothing
LV	Latent variable
M	Male
m	Multiplet
MAS	Meconium aspiration syndrome
Mat.	Maternal
MBB	Maternity Bissaya Barreto
MCADD	Medium chain acyl-coenzyme A dehydrogenase deficiency
MIR	Mid-infrared spectroscopy
MLEV-17	Malcolm Levitt- 17 sequence
MLPA	Multiplex Ligation Probe dependent Amplification
MRS	Magnetic resonances spectroscopy
MS	Mass spectrometry
MVA	Multivariate analysis
Mw	Molecular weight
NA	Not available

NaAD	Nicotinic acid adenine nucleotide
NAD ⁺	Nicotinamide adenine dinucleotide
NaDP	Nicotinicotinamide dinucleotide phosphate
NaMN	Nicotinic acid mono nucleotide
NLM	Non-linear mapping
NMND	<i>N</i> -methylnicotinamide
NMR	Nuclear magnetic resonance
NP	Non-pregnant
NS	Number of scans
nS-HR-MS	Nanospray high resolution mass spectrometry
NTD	Neural tube defect
OGTT	Oral glucose tolerance test
OPLS-DA	Orthogonal Partial Least Squares-Discriminant Analysis
OSC	Orthogonal signal correction
<i>p</i>	<i>p</i> -value
PAG	Phenylacetylglutamine
PAPP-A	Pregnancy associated plasma protein-A
PC	Principal Component
PCA	Principal Component Analysis
<i>p</i> -CS	<i>p</i> -cresol sulphate
<i>p</i> _{cvSE}	Covairance's standard error obtained by cross validation
PE	Preeclampsia
PG	Plasma glucose
PI	Phosphatidylinositol
PKU	Phenylketonuria
PLS-DA	Partial Least Squares-Discriminant Analysis
pPROM	Preterm premature ruptures of the membranes
PQN	Probabilistic quotient normalization
PROM	Premature rupture of the mebranes
PTD	Preterm delivery
<i>q</i>	Quartet
<i>Q</i> ²	Goodness of prediction or prediction power
QC	Quality control
Q-TOF	Quadropole time-of-flight
<i>r</i>	Correlation coefficient
<i>R</i> ² _x	Variation explained by the X matrix
<i>R</i> ² _y	Variation explained by the Y matrix
RDS	Respiratory distress syndrome
RF	Random forest
RF	Random forest
RG	Receiver gain
ROC	Receiver operating characteristic

RSPA	Recursive segment-wise peak alignment
RT	Retention time
s	Singlet
S/N	Signal to noise ratio
Sens	Sensitivity
SGA	Small for gestational age
SHY	Statistical Heterospectroscopy
SI	Size of real spectrum
SIMCA	Soft Independent Modeling of Class Analogy
SOM	Self-organizing maps
Spec	Specificity
SVM-RFE	Support Vector Machines-recursive feature elimination
SW	Spectral window
t	Triplet
T18	Trisomy 18
T1DM	Type 1 Diabetes Mellitus
T21	Trisomy 21
T2DM	Type 2 Diabetes Mellitus
TCA	Tricarboxylic acid
TD	Size of FID
THF	Tetrahydrofolate
TIC	Total ion chromatogram
TMA	Trimethylamine
TMAO	Trimethylamine- <i>N</i> -oxide
TN	True negative
TOCSY	Total Correlation Spectroscopy
TOF	Time-of-flight
TP	True positive
TPR	True positive rate
TSP	3-trimethylsilylpropionic acid
UA	University of Aveiro
uE3	Unconjugated estriol
Ui	Unassigned resonance i
UPCA	Unsupervised unfold principal component analysis
UPLC	Ultraperformance liquid chromatography
UV	Unit variance
VAST	Variable stability scaling
VD	Vaginal delivery
VIP	Variable importance for the projection
VIP _{cvSE}	VIP's standard error obtained by cross validation
VLBW	Very low birthweight
VLDL	Very low density lipoproteins

WDW	Type of window function
$Y_{pr[class]}$	Y predicted value reflective of sample class
β -hCG	β -human chorionic gonadotropin

1. Introduction

1.1 Healthy pregnancy and fetal development

Fetal development

Normal pregnancy has the duration of 38 to 40 gestational weeks (g.w.) and is divided into three trimesters, each of these periods characterized by specific hallmarks of fetal development (Moore *et al.*, 2008). The first trimester (1st T) corresponds to the period from fertilization to the 12th g.w. and it is during this period that organogenesis takes place and organ systems are most vulnerable. After fertilization the oocyte is formed and, by the end of the 2nd week, where implantation is completed. From that week onwards rapid anatomic development occurs and, by the 4th g.w., the complex vascular network of the placenta is formed enabling maternal-embryonic exchanges of gases, nutrients and metabolic wastes (Moore *et al.*, 2008, Atzori *et al.*, 2011, Kliegman *et al.*, 2011). Following this phase, the 4th to 8th g.w. are the most critical period in fetal development, as it is during this time that the main external and internal fetal structures begin to form (Moore *et al.*, 2008). Between week 9 to 12, primary ossification of the skeleton appears, limbs are partially formed, of the spleen produces red blood cells and urine production and discharge by the urethra into the amniotic fluid is initiated.

The following trimester (2nd T) corresponds to the period from the 12th to 26th g.w. and it is during this time that rapid growth and fetal organ maturation occurs. By the 24th g.w. the lungs begin to secrete surfactant to the developing alveoli, a necessary requirement for breathing after birth. The third trimester (3rd T) starts on the 26th g.w. and extends to the end of gestation, enhanced fetal growth occurring during this period. The fetus becomes able to synthesize its own body fat and fetal maturation is complete by week 37 (Moore *et al.*, 2008).

A representation of the most critical periods in fetal development is shown in Figure 1.1 highlighting the sensitive periods for adverse fetal development, and a timeline of development of specific anomalies and common sites of action of teratogens (Kliegman *et al.*, 2011). Through this figure it becomes clear that the most critical period in fetal development goes from the 3rd to 16th g.w. and that, during this period, major congenital anomalies can affect the central nervous system (CNS), heart, limbs, ears, eyes, teeth, palate and external genitalia.

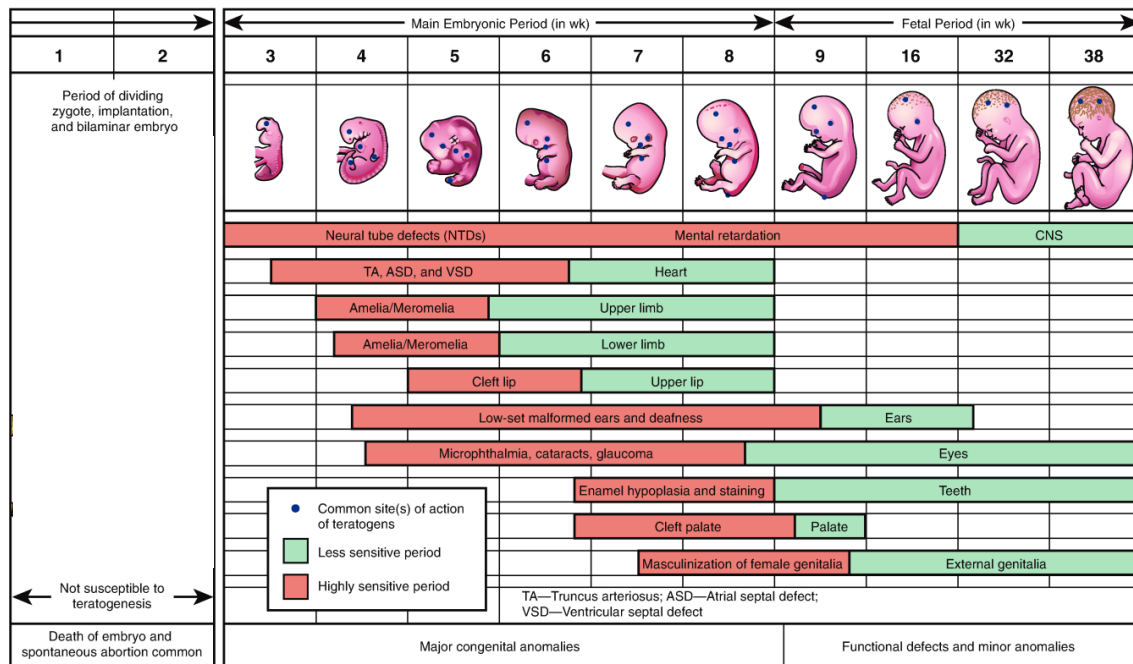


Figure 1.1: Representation of the critical periods in prenatal fetal development (adapted from (Kliegman *et al.*, 2011)).

Main maternal adaptations to pregnancy

During pregnancy, the mother undergoes profound physiological, endocrine and metabolic adaptations to 1) ensure the necessary requirements for the adequate growth and development of the fetus, 2) provide the fetus energy substrates required after birth, 3) ensure maternal needs to cope with the enhanced physiological demands of pregnancy and 4) provide the mother with sufficient energy stores and substrates necessary for labor and lactation (Hadden and McLaughlin, 2009). Maternal physiology is adjusted to support the requirements for fetal homeostasis and growth without compromising maternal well-being. These changes affect maternal cardiovascular, respiratory, renal and endocrine systems. Briefly, during pregnancy there is increased water and calcium retention, increased total blood volume to about 40%, increased cardiac output and heart rate. The respiratory changes are a consequence of the mechanical effect of the enlarging uterus that elevates the diaphragm position, of increased total oxygen consumption, mainly in the uterus and its contents, and of respiratory stimulant effects of progesterone, an important pregnancy hormone. Significant anatomic changes in the urinary collecting system occur due to progesterone action and enlarging uterus. Renal blood flow and glomerular filtration rate increase about 40% until mid-gestation due to increased maternal plasma volume. Endocrine adaptations include changes in insulin secretion, thyroid and adrenal hormone.

In pregnancy, insulin response to glucose stimulation is enhanced, and from early pregnancy until mid-gestation, fasting insulin levels are increased while reducing glucose levels. During this period, the anabolic effects of insulin are potentiated leading to inhibition of lipolysis, and gluconeogenesis whilst glycogen synthesis and storage in the liver is enhanced. Later in pregnancy, insulin resistance develops leading to impaired glucose tolerance possibly due to humoral factors such as increased human placental lactogen, cortisol and progesterone (Hacker *et al.*, 2004).

The placenta

The placenta and umbilical cord form a transport system for substances passing between mother and fetus (Moore *et al.*, 2008). The placenta plays a key role in maintaining the exquisite symbiosis between mother and fetus, as it is through this fetomaternal organ that all nutrient, gases and waste exchanges occur. The placenta is essential for protection, nutrition, respiration, excretion and hormone production, which are essential for maintaining pregnancy and promoting fetal development. Glycogen, cholesterol and fatty acids synthesis takes place in the placenta, which serves as nutrients and energy for the fetus, as well as pregnancy hormones human chorionic gonadotropin (hCG), progesterone and estrogen. The placenta also transports oxygen to the fetus and removes carbon dioxide and carbon monoxide from the fetus. In addition, it transports glucose, amino acids, vitamins, hormones, electrolytes and maternal antibodies while removing waste products such as urea and uric acid. It is also through the placenta that drugs, drug metabolites and infectious agents that may pose a teratogen risk (such as cytomegalovirus and *Toxoplasma gondii*) are transported (Hacker *et al.*, 2004, Moore *et al.*, 2008) .

Preparation for extrauterine life

Fetal adaptation to the extrauterine life begins before the onset of labor and is particularly challenged in premature newborns due to incomplete anatomic development and biochemical maturation (Moore *et al.*, 2008). Several days before labor, the fetal thyroid hormone dynamic is changed enabling the conversion of fetal thyroxine (T4) to triiodothyronine (T3), a metabolically more active form of the hormone and necessary for neonatal thermogenesis. Fetal breathing is also diminished with the onset of labor to decrease pulmonary fluid dynamics, an important step for respiration onset after delivery

and retention of surfactant in the lungs. The stress that labor imposes on the fetus also increases the release of catecholamines, responsible for glucose mobilization, lung fluid absorption, alterations in the perfusion of organs systems and possibly the onset of respiration (Hacker *et al.*, 2004). Shortly after birth, the overall status of the newborn is given by the Apgar score, a scoring system that allows the systematic assessment of the newborn's condition and immediate needs (Hacker *et al.*, 2004, Moore *et al.*, 2008). Apgar scores are assigned 1, 5 and 10 minutes after birth and are calculated by the sum of scores summarized in Table 1.1 for each minute. A normal score is considered 7 or higher in the 1st minute and 9 or higher in the 5th and 10th minutes (Hacker *et al.*, 2004).

Table 1.1: Apgar scoring for determining the newborns status (Hacker *et al.*, 2004).
^aresponse to foot sole stimulation.

Sign	Score		
	0	1	2
Heart rate	Absent	<100 beats/min	>100 beats/min
Respiratory effort	Absent	Slow, weak cry	Good, strong cry
Muscle tone	Limp	Some flexion of extremities	Active motion
Reflex irritability ^a	None	Grimace	Strong cry
Color	Pale, blue	Pink body, blue extremities	Completely pink

1.2 Prenatal disorders and diagnostic procedures: availability and needs

In spite of the remarkable adaptations that pregnant women undergo to ensure the healthy course of pregnancy and adequate fetal development, disruptions in this delicate maternal-fetal symbiosis may occur, which pose significant risk for the mother, fetus and newborn. Thus, adequate prenatal care and pregnancy monitoring are necessary to allow early intervention and improve fetal and maternal outcomes. Pregnancy disorders may affect mother, fetus or both, the main pathologies being summarized in Table 1.2, presented in alphabetical order, along with their prevalences, i.e. the percentage of pregnancies affected by each condition.

The work described in this thesis entailed the study of prenatal disorders with higher prevalences (due to the corresponding higher number of samples available), namely fetal malformations (FM), chromosomal disorders (CD), intrauterine growth restriction (IUGR), preeclampsia (PE), preterm delivery (PTD), gestational diabetes mellitus (GDM)

and premature rupture of the membranes (PROM). For other relevant disorders such as small for gestational age (SGA), fetal bleeding, intrauterine fetal demise (IUFD), placental abruption, placenta previa, oligohydramnios, polyhydramnios and uterine rupture, it was not possible to gather sufficient samples. A brief overview of each of these conditions (i.e. pathophysiology, diagnostic procedures and management) is presented throughout this section, in decreasing order of impact to the mother, fetus and/or newborn.

Table 1.2: Main prenatal disorders and corresponding prevalence. ADA: American Diabetes Association.

Prenatal disorder	Prevalence	Reference
Chromosomal disorders (CD)	0.5% of all live births	(Hacker <i>et al.</i> , 2004).
Fetal bleeding	0.1-0.8% of all pregnancies	(Hacker <i>et al.</i> , 2004)
Fetal Malformations (FM)	3-4% of all live births	(Bacino, 2012)
Gestational Diabetes Mellitus (GDM)	1-14% of all pregnancies	(ADA, 2004)
Intrauterine fetal demise (IUFD)	1% of all pregnancies	(Hacker <i>et al.</i> , 2004)
Intrauterine growth restriction (IUGR)	7-15% of all pregnancies	(Cetin and Alvino, 2009)
Oligohydramnios	1-5% of all pregnancies	(Kliegman <i>et al.</i> , 2011)
Placental abruption	0.5-1.5% of all pregnancies	(Hacker <i>et al.</i> , 2004)
Placental previa	0.5 of all pregnancies	(Hacker <i>et al.</i> , 2004)
Polyhydramnios	1-3% of all pregnancies	(Kliegman <i>et al.</i> , 2011)
Post-term pregnancy	6-12% of all pregnancies	(Hacker <i>et al.</i> , 2004)
Preeclampsia (PE)	5-8% of all live births	(Calhoun <i>et al.</i> , 2011)
Premature rupture of the membranes (PROM)	8-10% of all pregnancies	(Stuart <i>et al.</i> , 2005)
Preterm delivery (PTD)	9.6% of all live births	(Gracie <i>et al.</i> , 2011)
Small for gestational age (SGA)	< 10% of all live births	(Horgan <i>et al.</i> , 2010)
Uterine rupture	0.5 % of all pregnancies	(Hacker <i>et al.</i> , 2004)

1.2.1 Fetal Malformations

Fetal malformations (FM), also referred to as congenital anatomic anomalies, congenital malformations or birth defects, are fetal anatomic anomalies that develop during the fetal period and are present at birth. Malformations may be single or multiple, of varying severity (minor or major), affecting organs, tissues and limbs (Moore *et al.*, 2008). Some examples of fetal anomalies, involving the central nervous system, face, heart, lung,

abdominal wall, gastrointestinal tract, urinary system and skeletal, are summarized in Table 1.3.

Table 1.3: List of common fetal malformations (adapted from (Hacker *et al.*, 2004).

Affected organ or system	Malformation type
Central nervous system	Hydrocephaly, anencephaly, spina bifida
Face	Cleft lip and/or palate
Heart	Atrial septal defect, ventricular septal defect, tetralogy of Fallot, transposition of great vessels, arrhythmias
Lung	Congenital cystic adenomatoid malformations, lung sequestration, diaphragmatic hernia
Abdominal Wall	Gastrochisis, omphalocele
Gastrointestinal Tract	Bowel atresia or obstruction
Urinary System	Renal agenesis, polycystic kidney disease, hydronephrosis, posterior urethral valves
Skeletal	Skeletal dysplasia

Fetal malformations affect 3-4% of all live births, the most common type being malformations of the brain (10/1000 births), followed by malformations of the heart (8/1000 births), kidney (4/1000 births) and limbs (2/1000 births) (Moore *et al.*, 2008). The majority of malformations (about 50-60%) are of unknown etiology although some can be attributed to chromosomal disorders (described in detail in section 1.2.2), mutant genes, environmental agents (such as viruses or drugs) or multifactorial (genetic or environmental factors). Malformations originating from environmental factors account for 7-10% of malformations, with several teratogen effects of certain drugs and viruses being recognized. Thalidomide is an example of a teratogen drug, an anti-nausea drug used in the past, and now known to have caused abnormal development of limbs, namely meromelia (partial absence) and amelia (complete absence), facial anomalies, cardiac and kidney defects, while viral agents such as cytomegalovirus (CMV) causes microcephaly (small brain and skull), hepatosplenomegaly (enlarged liver and spleen), hydrocephaly (enlarged head), cerebral palsy (abnormal brain development causing impairment or total loss of motor function) and periventricular calcification in the brain. A summary list of malformations caused by thalidomide, CMV and other environmental agents is shown in Table 1.4.

Table 1.4: Some fetal malformations caused by environmental agents. FAS: fetal alcohol syndrome, IUGR: intrauterine growth restriction, CNS: central nervous system, NTD: neural tube defects, adapted from (Hacker *et al.*, 2004).

Drugs	Common congenital anomalies
Alcohol	FAS, IUGR, mental retardation, microcephaly, ocular anomalies, joint abnormalities, short palpebral fissures
Cocaine	IUGR, prematurity, microcephaly, cerebral infarction, urogenital anomalies, neurobehavioral disturbances
Methotrexate (anti-cancer)	Multiple anomalies, especially skeletal, involving the face, cranium, limbs, and vertebral column
Phenytoin (anti-convulsants)	Fetal hydantoin syndrome: IUGR, microcephaly, mental retardation, ridged frontal suture, inner epicanthal folds, eyelid ptosis, broad depressed nasal bridge, phalangeal hypoplasia
Thalidomide (anti-nausea)	Abnormal development of limbs, e.g., meromelia (partial absence) and amelia (complete absence), facial anomalies, systemic anomalies, e.g., cardiac and kidney defects
Valproic acid (anti-convulsant)	Craniofacial anomalies, NTDs, often hydrocephalus, heart and skeletal defects
Warfarin (anti-coagulant)	Nasal hypoplasia, stippled epiphyses, hypoplastic phalanges, eye anomalies, mental retardation
Infection agents	
Cytomegalovirus (CMV)	Microcephaly, chorioretinitis, sensorineural hearing loss, delayed psychomotor/mental development, hepatosplenomegaly, hydrocephaly, cerebral palsy, brain (periventricular) calcification
Human parvovirus B19	Eye defects, degenerative changes in fetal tissues
Rubella virus	IUGR, postnatal growth retardation, cardiac and great vessel abnormalities, microcephaly, sensorineural deafness, cataract, microphthalmos, glaucoma, pigmented retinopathy, mental retardation, newborn bleeding, hepatosplenomegaly, osteopathy, tooth defects
Toxoplasma gondii	Microcephaly, mental retardation, microphthalmia, hydrocephaly, chorioretinitis, cerebral calcifications, hearing loss, neurologic disturbances
Varicella virus	Cutaneous scars (dermatome distribution), neurologic anomalies, hydrocephaly, cataracts, microphthalmia, Horner syndrome, optic atrophy, nystagmus, chorioretinitis, microcephaly, mental retardation, skeletal anomalies (hypoplasia of limbs, fingers, and toes, etc.), urogenital anomalies

Diagnosis of fetal malformations

Ultrasound

The diagnosis of a malformation is carried out by ultrasound after the 16th weeks of gestation. In the 2nd T of gestation, the detection rates of these anomalies vary from 17 to 74%, depending on the population and the operator skill. In spite of the low sensitivity of the procedure (which corresponds to a low positive detection rate, i.e. low number of malformations correctly classified), it is very useful in detecting normal fetuses, with

specificity (true negative detection rate, i.e. the number of normal cases correctly classified) being around 100% (Hacker *et al.*, 2004).

Amniocentesis

Amniocentesis consists of removing amniotic fluid through the transabdominal insertion of a needle under ultrasound guidance. It is performed at 16-20 g.w. and has an associated 0.5% risk of pregnancy loss (Hacker *et al.*, 2004). In cases of equivocal or non-diagnostic fetal ultrasonic imaging, the determination of elevated alpha feto-protein (AFP, a fetal serum protein) levels in amniotic fluid allows the detection of neural tube defects, which in normal circumstances should be detectable only in trace amounts (Hacker *et al.*, 2004). Amniocentesis is also used in 3rd T to determine the risk of fetal lung maturity through the measurement of pulmonary phospholipids, used as lung surfactants, which enter the amniotic fluid from fetal lungs. Lung surfactants line lung alveoli diminishing the surface tension, thus lowering the pressure required to open alveoli and breath (Kliegman *et al.*, 2011). The presence of phosphatidylglycerol and a ratio of lecithin (or phosphatidylcholine) to sphingomyelin higher than 2.0 are indicators of normal lung surfactants levels, thus minimal risk of respiratory distress in the neonate (Hacker *et al.*, 2004).

1.2.2 Chromosomal Disorders

A normal human karyotype has 22 pairs of chromosomes, named autosomes, and a pair of sex chromosomes, i.e. females have two X chromosomes and males have one X and one Y chromosome. An example of a human male karyotype comprising 22 pairs of autosome chromosomes and an XY pair of sex chromosomes is shown in Figure 1.2. Chromosomal disorders (CD) are named numerical, in case there is loss or excess of chromosomes, or structural, if they involve rearrangements with or without loss or gain of the genetic material. CD affect 0.5% of all live births but it is estimated to affect 50% spontaneous abortions (Hacker *et al.*, 2004). Depending on the type of the CD, the affected chromosome and if it is balanced or unbalanced, the resulting phenotype (i.e. the individuals observable trait) is of variable severity.

Human male karyotype

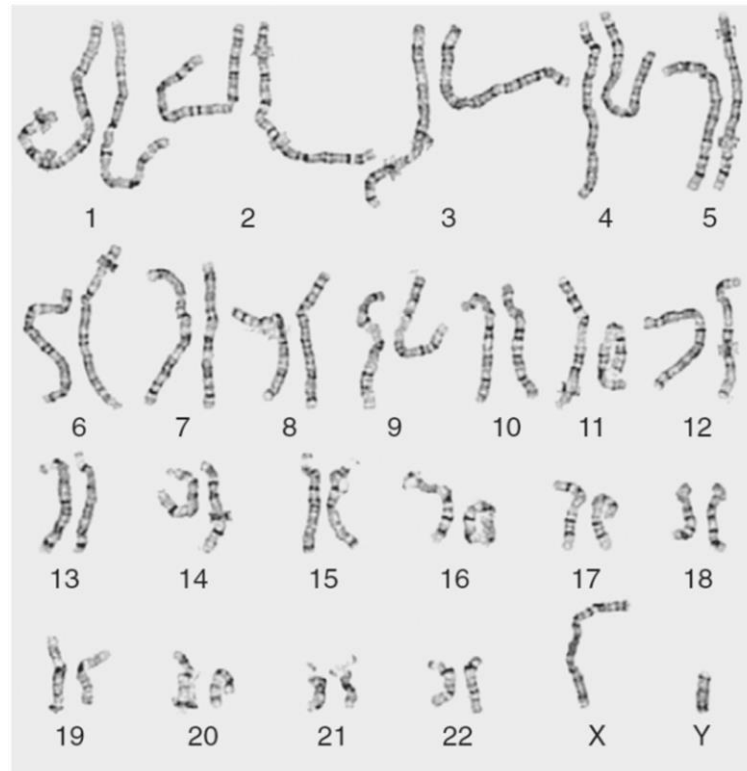


Figure 1.2: Human karyotype of a male subject (22 pairs of autosome chromosomes plus sex chromosomes pair XY). Adapted from (Kliegman *et al.*, 2011).

Numerical Chromosomal Disorders

Numerical CD, or aneuploidy, result from nondisjunction during gametogenesis, i.e. the failure of a chromosomal pair or pair of chromatids to disjoin during meiosis or mitosis (Moore *et al.*, 2008). These abnormalities can be monosomies, if there is only one chromosome of a given pair e.g. Turner's syndrome (45,X), trisomies, when there are three chromosomes of a given pair e.g. Down's syndrome (47,XX+21 or 47,XY+21), mosaicism, as is the case of two or more cell lines containing different number of chromosomes, tetrasomies, if there are four or five sex chromosomes and triploidies, when there is a complete extra set of chromosomes (Moore *et al.*, 2008).

Numerical disorders are unbalanced, since there is either gain or loss of the total genetic material, thus having a significant implication in the carrier's phenotype, affecting mental skills and/or causing anatomic defects with varying severity. Down syndrome (or trisomy 21, T21) is the most common numerical chromosomal disorder and subjects affected by this condition suffer from mental deficiency, brachycephaly, flat nasal bridge,

upward slant to palpebral fissures, protruding tongue, simian crease, clinodactyly of fifth digit, congenital heart defects and gastrointestinal tract abnormalities. A detailed list of some numerical disorders along with the corresponding karyotype, prevalence and phenotype is summarized in Table 1.5.

Table 1.5: List of types of chromosomal disorders, karyotypes, prevalence and phenotype adapted from (Moore *et al.*, 2008) and “Chromosome abnormalities” in www.genome.welcome.ac.uk.

Karyotype, syndrome and prevalence	Phenotype
45, X Turner Syndrome 2/10 000 female live births	Low height, sexual immaturity, short and thick neck, lymphoma, large thorax and cardiac anomalies.
47,XX+21 or 47,XY+21 Down Syndrome 15/10 000 births	Mental deficiency, brachycephaly, flat nasal bridge, upward slant to palpebral fissures, protruding tongue, simian crease, clinodactyly of fifth digit, congenital heart defects, gastrointestinal tract abnormalities.
47,XX+18 or 47,XY+18 Edward Syndrome 3/10 000 births	Mental deficiency, growth retardation, prominent occiput, short sternum, ventricular septal defect, micrognathia, low-set malformed ears, flexed digits, hypoplastic nails, rocker-bottom feet.
47,XX+13 or 47,XY+13 Patau Syndrome 2/10 000 live births	Mental deficiency, severe central nervous system malformations, sloping forehead, malformed ears, scalp defects, microphthalmia, bilateral cleft lip and/or palate, polydactyly, posterior prominence of the heels.
47, XXY Klinefelter Syndrome 2/10 000 male live births	Small testes, hyalinization of seminiferous tubules, aspermatogenesis, often tall with disproportionately long lower limbs. Intelligence is less than in normal siblings. Approximately 40% of these males have gynecomastia.
47, XXX 10/10 000 female live births	Normal appearance, usually fertile, 15-25% are mildly mentally retarded.
47, XYY 10/10 000 male live births	Normal appearance, usually tall, often exhibit aggressive behavior.

Structural Chromosomal Disorders

Structural CD can be balanced, if they involve rearrangements without loss or gain of the genetic material (e.g. translocations, inversions and insertions) or unbalanced if there is gain or loss of genetic material (e.g. deletions, duplications, ring chromosomes and isochromosomes (Moore *et al.*, 2008). A diagram of the formation of these chromosome structural abnormalities is shown in Figure 1.3.

Translocations occur when a part of a chromosome is transferred to the nonhomologous chromosome, as shown in Figure 1.3a while if the two nonhomologous chromosomes exchange parts it is called reciprocal translocation. Translocations are

commonly balanced thus do not necessarily cause an abnormal phenotype. Chromosome deletion occurs when there is a chromosome break with partial loss (Figure 1.3b), a specific type of deletions being ring chromosomes, when both ends of the chromosome are lost and rejoined forming a ring-shape chromosome (Figure 1.3c). Chromosome duplications refer to the duplication of part of a chromosome, and they can occur within a chromosome, attached to a chromosome or as a separate fragment (Figure 1.3d). Inversions are chromosomal aberrations where a segment of the chromosome is inverted as shown in Figure 1.3e. Inversions can be paracentric, if only one arm of the chromosome is involved, or pericentric, if the both arms and the centromere are involved. If the centromere division occurs transversely instead of longitudinally, i.e. with loss of one arm and duplication of the other arm, it is called isochromosome (Figure 1.3f), this being the most common type of structural abnormality of the X chromosome (Moore *et al.*, 2008).

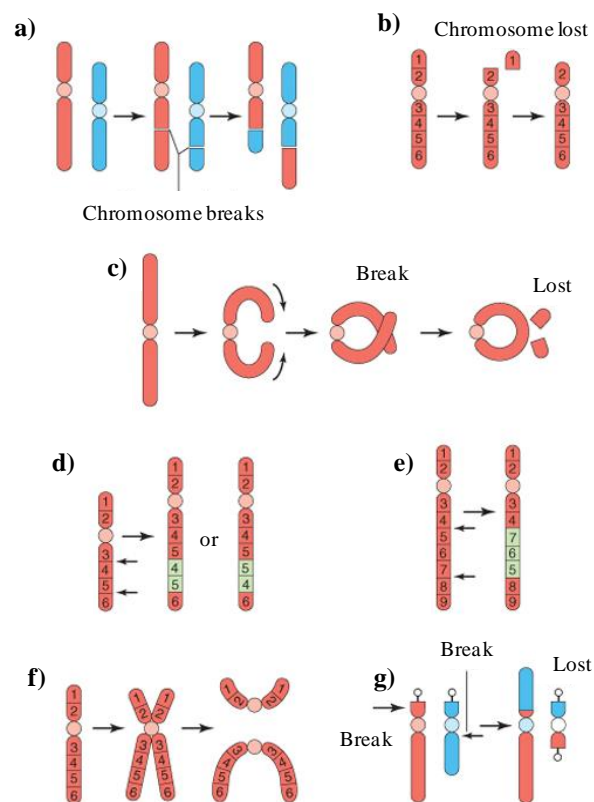


Figure 1.3: Diagrams illustrating various structural chromosomal abnormalities: a) reciprocal translocation, b) terminal deletion, c) ring chromosome, d) duplication, e) Paracentric inversion, f) Isochromosome, g) Robertsonian translocation. Arrows indicate how the structural abnormalities are produced. Adapted from (Moore *et al.*, 2008).

Specific chromosome deletions and duplications are associated with certain syndromes that pose significant phenotypic implications. An interesting example is Prader-Willi and Angelmans syndromes, which are caused by the loss of the critical region involving bands 1 to 3 of the proximal region of the long arm chromosome 15 del15(q12q13). If the deletion occurs in the paternal chromosome the offspring will have Prader-Willi syndrome, whilst if it occurs in the maternal chromosome the offspring will have Angelmans syndrome, characterized by microcephaly, macrosomia, ataxia, excessive laughter, seizures, severe mental retardation (Moore *et al.*, 2008). These and other syndromes are summarized in Table 1.6 along with the related chromosome abnormality, prevalence (when available) and phenotype.

Table 1.6: Examples of genetic syndromes, chromosomal abnormality, prevalence (when available) and phenotype. q: long arms and p: short arm of a chromosome. Adapted from (Struthers *et al.*, 2002, Tharapel *et al.*, 2002, Clayton-Smith and Laan, 2003, Moore *et al.*, 2008, Shprintzen, 2008, Morris, 2010, Weksberg *et al.*, 2010).

Syndrome, chromosome abnormality and prevalence	Phenotype
Prader-Willi Syndrome del15q12 (most cases) 1/16 000 live births	Hypotonia, hypogonadism, obesity with hyperphagia, distinct face, short stature, small hands and feet, mild developmental delay. Paternal origin.
Angelman's Syndrome del15q12 (most cases) 1/10 000-40 000 live births	Microcephaly, macrosomia, ataxia, excessive laughter, seizures, severe mental retardation. Maternal origin.
Miller-Dieker Syndrome del17p13.3 (most cases)	Type I lissencephaly, dysmorphic face, seizures, severe developmental delay, cardiac anomalies.
DiGeorge Syndrome del22q11 (some cases) 1/4000 live births	Thymic hypoplasia, parathyroid hypoplasia, conotruncal cardiac defects, facial dysmorphism.
Velocardiofacial Syndrome del22q11 (most cases) 1/2 000 live births	Palatal defects, hypoplastic alae nasi, long nose, conotruncal cardiac defects, speech delay, learning disorder, schizophrenia-like disorder.
Smith-Magenis Syndrome del17p11.2 1/25000-50000 live births	Brachycephaly, broad nasal bridge, prominent jaw, short broad hands, speech delay, mental retardation.
Williams Syndrome del17q11.23 (most cases) 1/7 500 live births	Short stature, hypercalcemia, cardiac anomalies, especially supraaortic stenosis, characteristic elfin-like face, mental retardation.
Beckwith-Wiedemann Syndrome dup11p15 (some cases) 1/13 700 live births	Macrosomia, macroglossia, omphalocele (some cases), hypoglycemia, hemihypertrophy, transverse ear lobes.

Diagnosis of chromosomal disorders

The probability of carrying a fetus affected by an aneuploidy significantly increases with maternal age, the estimated risk of having a child with trisomy 21, 18 and 13 of 1 in 1000, 1 in 2500 and 1 in 8000 respectively at age 20 whilst at age 35 the risk increases to 1 in 250, 1 in 600 and 1 in 1800 (Nicolaidis, 2011). Conversely, for other chromosomal disorders such as Turner syndrome (45,X), abnormalities in sex chromosomes (47,XXX, 47,XXY and 47,XYY), and triploidies (defined as an entire extra set of chromosomes, i.e. 69,XXX or 69,XXY or 69,XYY) have no connection with maternal age (Nicolaidis, 2011). Until recently, the gold standard for detecting fetal chromosomal disorders was amniocentesis performed at the 2nd T. Recently, non-invasive tests performed at the 1st T, allow the earlier suspicion of CD through the measurement of maternal serum biomarkers, alpha-fetoprotein (AFP), free β -human chorionic gonadotropin (β -hCG), inhibin A, unconjugated estriol (uE3) and pregnancy associated plasma protein-A (PAPP-A), combined with ultrasound measurement of the nuchal translucency (Hourrier *et al.*, 2010). Definitive confirmation, in the 1st T, requires chorionic villus sampling (CVS) or amniocentesis in the 2nd T of pregnancy. These diagnostic procedures will be described in more detail below.

Amniocentesis, Chorionic villus sampling (CVS) and Cordocentesis

After retrieving amniotic fluid through amniocentesis, amniotic cells require approximately 2 weeks of culture before fetal karyotyping can be performed, although fluorescent insitu hybridization (FISH) and Multiplex Ligation Probe dependent Amplification (MLPA) can give preliminary results for specific chromosome abnormalities, i.e. in chromosome 21, 13 and 18 within 24-48 hours (Scott *et al.*, 2003, Hacker *et al.*, 2004). Other genetic disorders, such as Prader-Willi syndrome, Angelman syndrome, DiGeorge syndrome and Wiliam syndrome, and also cystic fibrosis, Tay-Sachs disease, can be diagnosed via amniocentesis by amplifying fetal DNA in amniocytes using a polymerase chain reaction (Hacker *et al.*, 2004). CVS consists of removing a small sample of chorionic villi transabdominally or transcervically under ultrasound guidance with an associated fetal loss rate inferior to 1% (Hacker *et al.*, 2004). Chorionic villi are branches emerging from the chorion, on the fetal portion of the placenta that maximizes the contact surface with maternal blood, as shown in Figure 1.4 (Moore *et al.*, 2008). CVS

is usually performed from the 9th to 12th gestational weeks (before amniocentesis could be performed), enabling earlier diagnosis, although with diagnostic precision somewhat inferior to that of amniocentesis due to 1% risk of mosaicism caused by placental contamination (Scott *et al.*, 2003, Hacker *et al.*, 2004). Fetal karyotyping is performed after tissue culture and results are obtainable in 6 to 8 days (Hacker *et al.*, 2004).

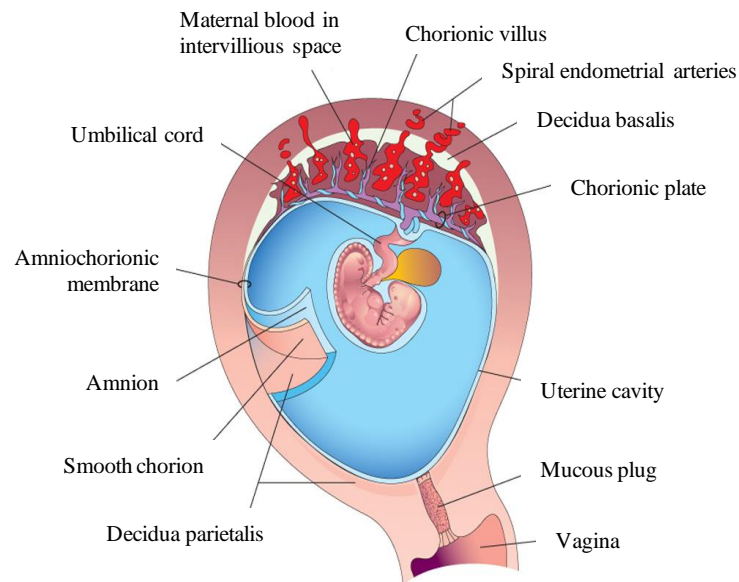


Figure 1.4: Graphic representation of a gravid uterus at the 4th week of gestation showing fetal membranes and embryo, adapted from (Moore *et al.*, 2008).

Cordocentesis, or fetal blood sampling, consists of transabdominally removing fetal blood directly from the umbilical vein under ultrasound guidance. The procedure is performed after the 17th g.w. and has an associated pregnancy loss rate of 1%. Because fetal leucocytes are more rapidly cultured than amniocytes, fetal karyotyping is faster through than amniocentesis, the result being available in 3 days (Scott *et al.*, 2003, Hacker *et al.*, 2004).

Maternal serum markers

In the presence of a fetal aneuploidy some feto-placental products levels, such as alpha-fetoprotein (AFP), free β -hCG, inhibin A, uE3 and PAPP-A are changed in maternal serum (Nicolaidis, 2011). The measurement of these maternal serum markers has a detection rate for T21 (Down syndrome) at the 2nd trimester of pregnancy that varies from 60 to 65% when combined with AFP and free β -hCG with maternal age, 65-70% with the addition of uE₃, and 70-75% with the addition of inhibin A, at a false positive rate of 5%

(Nicolaides, 2011). In the 1st T, T21 may be detected by the measurement of maternal serum AFP and free β -hCG in combination with ultrasound, as fetuses with trisomy 21 have increased nuchal translucency, absence of the nasal bone, increased impedance to flow in ductus venous and tricuspid regurgitation (Hacker *et al.*, 2004). When combining nuchal translucency with maternal age, and maternal serum free β -hCG and PAPP-A, named triple test, the detection rate for trisomy 21 increases to 93-96%, with a 2.5% false positive rate (Nicolaides, 2011). Other chromosomal disorders, such as trisomy 18 and 13 are associated with decreased maternal serum free β -hCG and PAPP-A, whilst abnormalities in sex chromosome are associated with decreased PAPP-A and normal free β -hCG (Nicolaides, 2011).

1.2.3 Intrauterine Growth Restriction

Intrauterine growth restriction (IUGR) is defined as fetal growth below the 5th percentile being growth restricted fetuses at higher risk of meconium aspiration, asphyxia, polycythemia, hypoglycemia and mental retardation. IUGR differs from small for gestational age (or SGA) as the latter refers to cases with fetal growth below a given reference range whilst in IUGR fetal growth potential is restricted by a pathological process in utero (Scott *et al.*, 2003, Hacker *et al.*, 2004). The condition affects 7 to 15% of all pregnancies (Cetin and Alvino, 2009), the origin of the disorder being maternal, placental or fetal, although a combination of these is frequently found in IUGR affected pregnancies (Hacker *et al.*, 2004). Some causes of IUGR include poor nutritional intake, smoking, drug abuse, alcoholism, cyanotic heart disease and pulmonary insufficiency, whilst other conditions that lead to inadequate placental transfer of nutrients can also cause IUGR e.g. hypertension, chronic renal disease and pregnancy-induced hypertension, or situations leading to inadequate substrates levels in the fetal compartment, such as intrauterine infection (toxoplasmosis, CMV, rubella and herpes simplex) and congenital malformations (Hacker *et al.*, 2004). IUGR can be symmetric, in case the head and body are equally restricted and the head to abdominal ratio is normal, although at an inferior growth rate, or asymmetric, usually developed in late pregnancy, if brain growth is preferred over other organs resulting in a head size proportionally larger than the abdominal size (Hacker *et al.*, 2004).

Diagnosis of IUGR

IUGR is diagnosed by ultrasound, being 50-90% of all cases detectable through this method (Hacker *et al.*, 2004). The assessment of several sonographic parameters is used to diagnose IUGR, i.e. biparietal diameter, head circumference, head to abdominal circumference ratio, femoral length, femoral length to abdominal circumference ratio, amniotic fluid volume, calculated fetal weight and abdominal circumference, the latter being the single most effective predictor of fetal weight as it is reduced in both symmetric and asymmetric IUGR (Hacker *et al.*, 2004). Apart from the anatomic changes in IUGR fetuses, abnormalities in the placental vascular tree are known to cause variations in fetal circulation that can be detected through Doppler velocimetry of the umbilical artery (Scott *et al.*, 2003).

1.2.4 Preeclampsia

Preeclampsia (PE) is defined by the new onset of the triad hypertension, proteinuria and generalized edema (Scott *et al.*, 2003, Hacker *et al.*, 2004), usually after the 20th week of gestation, and affects 5-8% of all live births (Calhoun *et al.*, 2011). It usually affects nulliparas although in multiparas with twins and chronic hypertension are at increased risk of developing this condition (Hacker *et al.*, 2004). PE may be present in mild or severe forms, depending on the severity of hypertension and proteinuria and the degree of other organs involvement (Scott *et al.*, 2003, Hacker *et al.*, 2004). A variant of severe PE is the Hemolysis, Elevated Liver enzymes and Low Platelets (HELLP) syndrome, more common in multiparous women, older than 25 and prior to 36 g.w.. Another form of severe PE includes the occurrence of seizures unrelated to other causes (Scott *et al.*, 2003, Hacker *et al.*, 2004). The condition poses significant risk for the mother, such as development of disseminated intravascular coagulation, intracranial hemorrhage, renal failure, retinal detachment, pulmonary edema, liver rupture, placental abruption and death, whilst neonatal outcome depends to gestational age at delivery (Scott *et al.*, 2003). PE is characterized by vasospasm, hemoconcentration and ischemic changes in the placental, kidney, liver and brain, with several causative mechanisms having been suggested, i.e. placental, immunologic, endocrine, nutritional, infectious agents and genetic predisposition origins, the exact etiology remaining unclear (Scott *et al.*, 2003, Hacker *et al.*, 2004).

Diagnosis and Management of PE

PE diagnosis is mainly based on the development of hypertension, i.e. systolic blood pressure higher than 140 mm Hg and diastolic blood pressure higher than 90 mm Hg, in a woman with previous normal blood pressure, and 24 hour urine proteinuria higher than 300 mg. The condition is often accompanied or preceded by generalized edema, mainly in the lower extremities but also in the face, hands or lungs (Scott *et al.*, 2003, Hacker *et al.*, 2004). Delivery is the only definitive resolution, thus PE management depends on the gestational age at diagnosis and must be a compromise between preventing maternal complication and minimizing neonatal morbidity arising from prematurity (Scott *et al.*, 2003, Hacker *et al.*, 2004). Currently, there are no predictive or preventive methods for PE and, given the disorders' significant impact for mother and newborn, the development of reliable methods are of paramount importance (Scott *et al.*, 2003).

1.2.5 Preterm Delivery

Preterm delivery (PTD), or preterm labor, is defined as birth occurring between 20th and 37th gestational weeks. PTD occurs in 9.6% of all live births and it is associated with 50 to 70% of neonatal morbidity and mortality (Hacker *et al.*, 2004). PTD can be classified into three subtypes, spontaneous PTD or idiopathic, preterm premature ruptures of the membranes (PPROM or pPROM) or induced labor due to medical indication. The possibility of identifying women at risk of a preterm delivery in order to allow the effective intervention and/or prevention of PTD or to decrease neonatal morbidity and mortality is a goal with utmost relevance. Although some findings have been associated with the risk of developing PTD although no reliable predictive method is yet available. Cervico-vaginal infections, such as bacterial vaginosis, and short cervical length have been associated with the occurrence of PTD. Cervical length is measured by transvaginal ultrasound and women with cervical length inferior to 2.5 cm at 24-28 g.w. have a 3- to 5-fold higher risk of PTD. At the same time of gestation, the presence of fetal fibronectin, a protein produced in fetal membranes, in the cervix and vagina predicts more than half of the spontaneous preterm births before the 28 g.w., while its absence is associated with low risk of PTD (Hacker *et al.*, 2004). Salivary estriol, a fetal product increased prior to spontaneous birth, has also

been tested for prediction of PTD although with low sensitivity and specificity and high false positive rates (71%, 77% and 23%, respectively) (Scott *et al.*, 2003).

Diagnosis and Management of PTD

The diagnosis of PTD follows two criteria: a) uterine contractions (>4 per 20 minutes or 8 per 60 minutes) and cervical dilation (2 cm in a nullipara, a woman who has never given birth, or 3 cm in a multipara, a woman who has given birth) and cervical effacement (cervical thinning expressed in percentage) higher than 80% or b) uterine contractions and cervical change, the latter being the most well accepted criteria although vulnerable to bias. After PTD diagnosis, women are given adequate hydration and bed rest and uterine contractions cease in approximately 20% of the cases. When the previous attempt is not successful uterine tocolytic therapy (anti contraction medications) is initiated, namely with magnesium sulphate, nifedipine and prostaglandin synthetase inhibitors as well as antibiotics to prevent intra-amniotic and fetal infections. Furthermore corticoid therapy, specifically bethamethasone, is given to promote fetal lung maturation, reducing the incidence of respiratory distress syndrome (or RDS) and mortality in the newborn (Hacker *et al.*, 2004).

1.2.6 Gestational Diabetes Mellitus

Gestational diabetes Mellitus (GDM) is defined as the development of carbohydrate intolerance with onset or first recognition during pregnancy. The condition varies in terms of severity and is known to affect 1 to 14% of all pregnancies (ADA, 2004, Hod *et al.*, 2008). In terms of pathophysiology, GDM is similar to that of Type 2 Diabetes Mellitus (T2DM) as it involves pancreatic beta-cell dysfunction leading to impaired insulin secretion and, subsequently, increased insulin resistance (ADA, 2004, Hod *et al.*, 2008, Creasy *et al.*, 2009). The occurrence of GDM often increases the risk of adverse pregnancy outcomes, such as fetal macrosomia, neonatal hypoglycemia, perinatal mortality, jaundice, polycythemia, hypocalcemia, development of maternal hypertensive disorders, need for cesarian delivery and, in more severe cases, fetal demise in the last 4-8 weeks of gestation (ADA, 2004, Hod *et al.*, 2008). GDM may also pose long-term complications for the mother and child such as post-natal maternal T2DM and increased risk of obesity, glucose

intolerance, T2DM and cardiovascular disease in the offspring (ADA, 2004, Hod *et al.*, 2008). In GDM-affected pregnancies, maternal metabolism is perturbed, increased insulin resistance and β -cell dysfunction being known to cause deregulation in carbohydrates, lipids and amino acids metabolism (Hod *et al.*, 2008). Decreased insulin secretion leads to decreased glucose uptake by insulin-dependent tissues causing maternal hyperglycemia. Because glucose transfer to the fetus is concentration-dependent, maternal hyperglycemia leads to increased placental transfer of glucose and subsequently fetal hyperglycemia and hyperinsulinism, a condition leading to the development of macrosomia. As fetal hyperinsulinism remains after delivery, the sudden disruption of umbilical cord glucose supply to the newborn increases the risk of neonatal hypoglycemia, a condition that can lead to brain damage in the newborn (Hod *et al.*, 2008). GDM is associated with increased levels of triglycerides and cholesterol, further increasing the hyperlipidemia condition occurring in pregnancy, and also increasing the risk of fetal macrosomia. The origin of hyperlipidemia has been associated with hormonal dysfunction, and GDM has been found to increase placental lipids fluxes and enhanced fatty acids consumption by the fetus in GDM affected pregnancies (Hod *et al.*, 2008, Herrera and Ortega-Senovilla, 2010). GDM is also associated with increased oxidative stress, characterized by placental endothelial dysfunction, increased formation of reactive species and decreased anti-oxidant defenses whereas macrosomic infants from GDM mothers have lower antioxidant defenses, and higher products of oxidation reactions (Hod *et al.*, 2008, Herrera and Ortega-Senovilla, 2010).

Diagnosis and Management of GDM

GDM screening is usually performed at 24-28 g.w. initially by a 50 g glucose challenge followed by the measurement of plasma glucose (PG) after 1 hour named O'Sullivan test. If PG is higher than 140 mg/dL (7.8 mmol/L) a subsequent 100 g or 75 g oral glucose tolerance test (OGTT) should be performed, the latter being the current recommendation by the American Diabetes Association- ADA (Scott *et al.*, 2003, Hacker *et al.*, 2004, ADA, 2013). After the glucose challenge plasma glucose (PG) is measured at fasting (FPG) and after 1 and 2 hours, for the 75 g challenge and at fasting and after 1, 2 and 3 hours, for the 100 g challenge. GDM is then diagnosed if two or more values are above the cutoffs for the 75 g challenge or if three or more values are above the cutoff for

the 100 g challenge. The cutoff values for O’Sullivan, and OGTT with 75 g and 100 g are summarized in Table 1.7.

Table 1.7: Criteria for the diagnosis of GDM for O’Sullivan, 100 g oral glucose tolerance test (OGTT) and 75 g OGTT, adapted from (Hod *et al.*, 2008).

	O’Sullivan 50 g test		100 g OGTT		75 g OGTT	
	mg/dL	mmol/dL	mg/dL	mmol/dL	mg/dL	mmol/dL
Fasting	—	—	95	5.3	95	5.3
1 hour	140	7.8	180	10.0	180	10.0
2 hours	—	—	155	8.6	155	8.6
3 hours	—	—	140	7.8	—	—

GDM management aims to achieve euglycemia (normal blood glycemic levels) through diet, insulin therapy and exercise, in order to decrease the occurrence of GDM complications. In the dietary intervention caloric intake is calculated with basis on the ideal body weight and is normally efficient for the management of GDM patients with normal fasting PG (FPG). Patients with FPG higher than 105 mg/dL or 2 hour post prandial PG higher than 120 mg/dL need to complement the dietary intervention with insulin therapy. Other oral hypoglycemic agents are not recommended in pregnancy as they can cross the placenta and pose risk of teratogenesis and neonatal hypoglycemia. Furthermore, GDM patients are encouraged to exercise for about half an hour after meals (Hacker *et al.*, 2004).

1.2.7 Premature Rupture of the Membranes

Premature rupture of the membranes (PROM) is defined as the spontaneous rupture of membranes prior to the onset of labor at any stage of gestation. It is the most common pregnancy disorder, affecting 8-10% (Stuart *et al.*, 2005) of all pregnancies and 30% of preterm deliveries (Scott *et al.*, 2003). The etiology of PROM remains unclear although some conditions such as cervical incompetence, polyhydramnios (increased amniotic fluid), cervical and vagina infections, abnormal membrane physiology and nutritional deficiency are thought to contribute to its occurrence (Scott *et al.*, 2003, Hacker *et al.*, 2004). The occurrence of PROM at term is associated with increased risk of maternal and neonatal sepsis and the need for neonatal resuscitation after delivery (Stuart *et al.*, 2005).

Diagnosis and Management of PROM

The diagnosis of PROM is based on vaginal loss of amniotic fluid, confirmed by testing for the presence of amniotic fluid with nitrazine papers which turns blue in the presence of alkaline amniotic fluid or by microscopic observation of arborization or “ferning” of dried amniotic fluid. These methods have high accuracy (93% and 96% for nitrazine and microscopic evaluation respectively) although high false positive rates are common due to contaminations with blood, semen, alkaline urine, bacterial vaginosis, trichomoniasis (Scott *et al.*, 2003, Hacker *et al.*, 2004). As the amniotic sac serves as a barrier against infection, PROM poses higher risk of infection and sepsis. Thus, PROM management depends on the gestational age at the time of membrane rupture. If PROM occurs after 36th weeks of gestation, labor should be induced after 6-12 hours in the absence of spontaneous contractions are absent. If PROM occurs prior to the 36th gestational week infection *vs.* prematurity must be weighted as the risk of infection increases with the duration of membrane rupture. Antibiotic therapy is recommended if chorioamnionitis, or intraamniotic infection, is confirmed and labor must be induced once antibiotics are started, whilst in the absence of infection tocolytic therapy can be useful in prolonging pregnancy (Scott *et al.*, 2003, Hacker *et al.*, 2004).

In spite of the plethora of diagnostic procedures used for prenatal care, there is a continuous need to improve currently available methods, and to do this by developing reliable non-invasive tests that pose no harm to the mother and developing fetus. Moreover, there is a recognized lack of predictive methods for disorders developed in late pregnancy such as IUGR, PE, PTD and GDM, conditions associated with significant complications for mother, fetus and newborn. Thus, metabolomics emerges as valuable methodology to identify potential markers in connection with pregnancy disorders. The basic concepts underlying a metabolomic study are presented in the following section.

1.3 Metabolomics: envisaging new clinical strategies in prenatal health

The terms metabolomic and metabonomic are often used interchangeably, in spite of slight differences in their history and definitions. Metabolomics is defined as the comprehensive measurement and quantification of all metabolites in a living system (Fiehn, 2002), whilst metabonomics has been defined as “the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or

genetic modification”, thus including the concept of metabolome perturbation and need for advanced statistical methods to detect the perturbation effects (Nicholson *et al.*, 1999). Metabolomics is one of the “omics” techniques together with genomics, transcriptomics and proteomics that study the genome (DNA), the transcriptome (RNA) and proteome (proteins), respectively. The term genomics encompasses the study of the set of genes describing an organism, the genome, whilst transcriptomics reflects the study of changes in gene expression which are then translated into proteins, thus emerging proteomics as the study of proteins in a living system (Figure 1.5). However, the complex interaction between genes and environment, such as lifestyle and diet, induce changes in biochemical processes which can be difficult to detect with proteomics or transcriptomics. On the other hand, such interactions are known to influence the metabolome very significantly. Indeed, it has been realized that small molecules, or metabolites (low molecular weight compounds, $M_w < 1500$ Da), involved in biological processes provide a great deal of

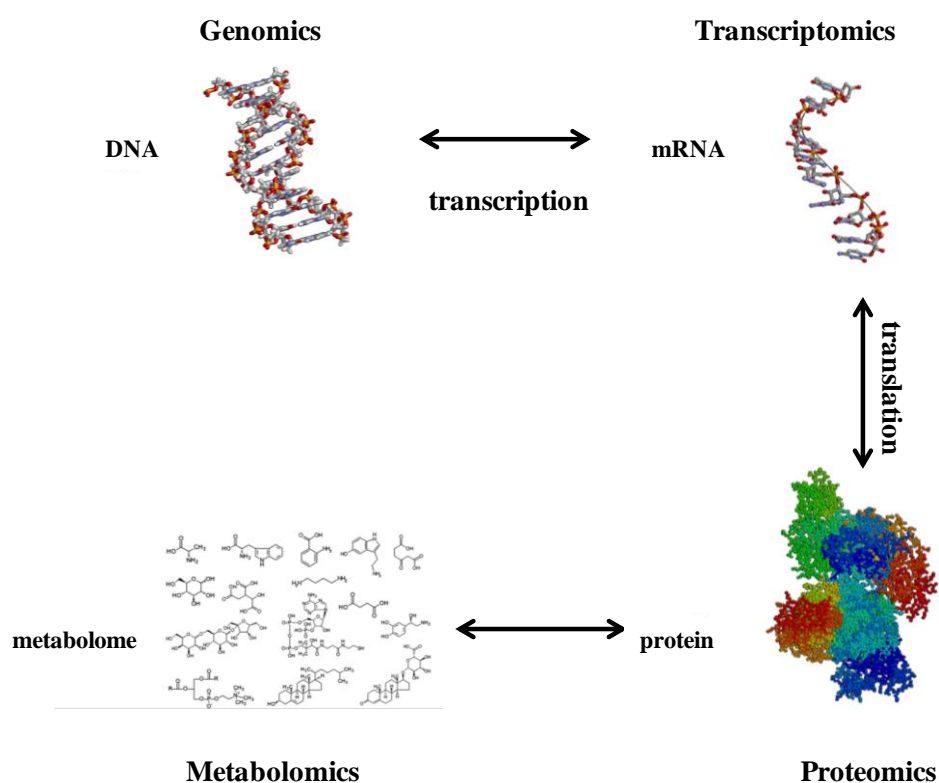


Figure 1.5: Schematic representation of the “Omic” technologies showing information flow from genomics to transcriptomics, proteomics and metabolomics (adapted from (Dunn *et al.*, 2011, Roux *et al.*, 2011)).

information of the whole organism and its interactions with both internal and external perturbations (Lindon *et al.*, 2007, Roux *et al.*, 2011).

The value of metabolomics has been well recognized in the field of molecular toxicology, pharmacology and disease research (Lindon *et al.*, 2004, Lindon *et al.*, 2006, Lindon *et al.*, 2007, Coen *et al.*, 2008, Duarte *et al.*, 2013). These studies are pursued through the analysis of biofluids, such as urine, blood, cerebrospinal fluid, amniotic fluid, saliva, or intact tissues, tissue extracts, cell cultures or in vitro cell systems (Lindon *et al.*, 2000, Lindon *et al.*, 2007). In the context of disease research, the mostly used biofluids are urine and blood plasma since they are easily collectable and hence can easily be used for disease diagnosis and monitoring but also for clinical trials and drug therapy monitoring (Lindon *et al.*, 2004, Lindon *et al.*, 2007). Some of the diseases more extensively studied through metabolomics comprise cancer (Madsen *et al.*, 2009, Rocha *et al.*, 2010, DeFeo *et al.*, 2011, Duarte and Gil, 2012), diabetes (Salek *et al.*, 2007, Makinen *et al.*, 2008, Madsen *et al.*, 2009, Friedrich, 2012), cardiovascular diseases (Madsen *et al.*, 2009, Griffin *et al.*, 2011, Wurtz *et al.*, 2011, Senn *et al.*, 2012), neurological diseases (Sussulini *et al.*, 2009, Lutz and Cozzzone, 2011, Cai *et al.*, 2012, Hassan-Smith *et al.*, 2012, Xu *et al.*, 2012, Zheng *et al.*, 2013), respiratory diseases (Slupsky *et al.*, 2009, Saude *et al.*, 2011), gastroenterological diseases (Bernini *et al.*, 2011, Schicho *et al.*, 2012) among others (Madsen *et al.*, 2009, Emwas *et al.*, 2013). In disease research, the application of metabolomic has evolved towards the use of large-scale epidemiological data to unveil not only the disease-related profiles but also the profile associated with disease risk factors, for instance high blood pressure as a major risk of coronary heart disease and stroke (Holmes *et al.*, 2008). The potential of metabolomics has also been tested for disease prediction, for instance type 1 and type 2 diabetes mellitus (Oresic *et al.*, 2008, Wang *et al.*, 2011) and disease prognosis (Makinen *et al.*, 2008, Nicholson *et al.*, 2012). Another emerging concept is that of pharmacometabolomics which aims to predict the effect of drug administration prior to dosing, towards the implementation of personalized medicine procedures (Nicholson *et al.*, 2011, Nicholson *et al.*, 2012). This hypothesis has been tested to predict acetaminophen (a common analgesic) metabolism (Clayton *et al.*, 2009), to predict anti-diabetic drugs response (Qiu *et al.*, 2008), and the metabolic effects of chemotherapy agents (Keun *et al.*, 2009).

A general scheme of a metabolomic study based on NMR and MS is shown in Figure 1.6, showing a typical workflow from data acquisition to MVA, possible identification of biomarkers of the condition being studied, which can be ultimately used to elaborate a biological outcome. Metabolomic studies are strongly based in Nuclear Magnetic Resonance (NMR) and Mass Spectrometry (MS) techniques, their principles being described in the following section 1.4, although Infrared (IR) spectroscopy has also been used occasionally. NMR and MS are suitable for the study of biofluids since they provide large amounts of potential information on composition of complex mixtures, in one single run. The use of proton (^1H) NMR is advantageous due to its robustness, accuracy and reproducibility and it is well recognized for the study of complex mixtures.

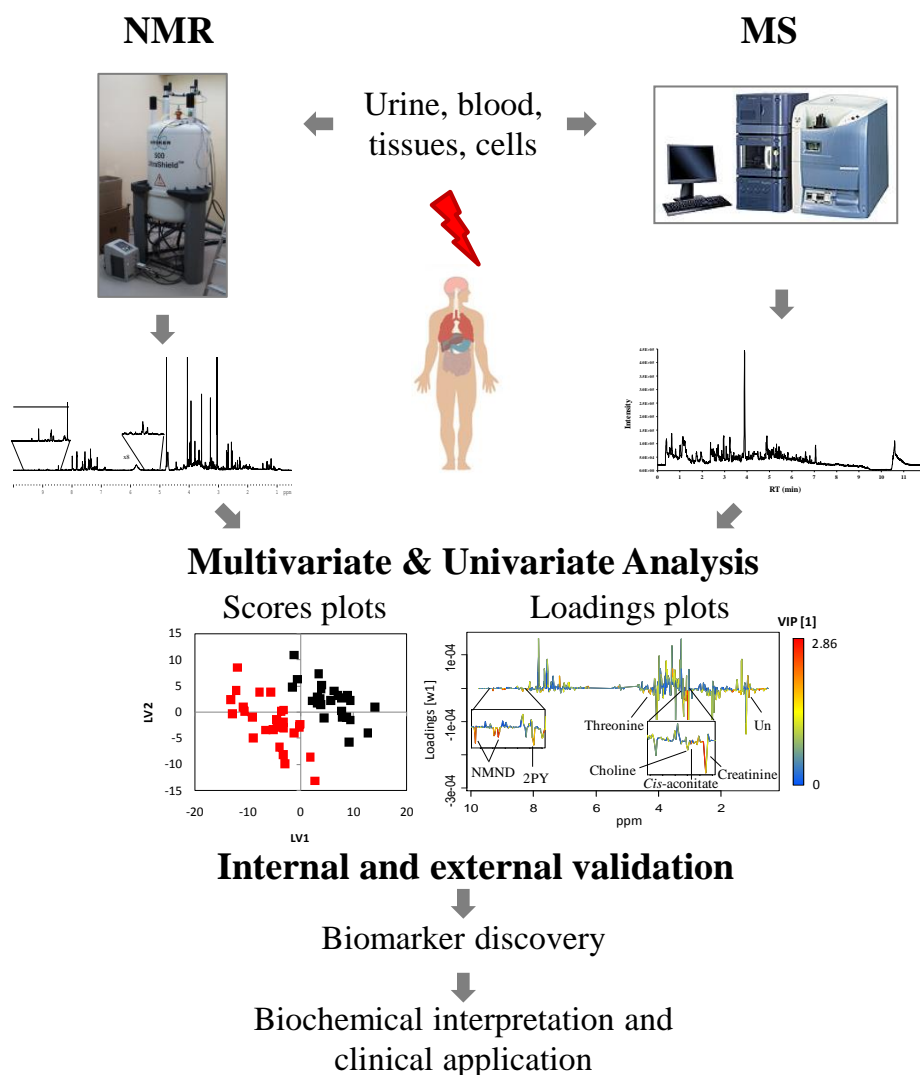


Figure 1.6: General scheme of a metabolomic study using NMR and MS methods.

However, NMR lacks sensitivity which may be complemented through the use of MS platforms, the latter technique requiring a pre-separation step, usually liquid chromatography (LC), gas chromatography (GC), or more recently ultraperformance LC (UPLC) (Lindon *et al.*, 2000, Lindon *et al.*, 2004, Lenz and Wilson, 2007, Lindon *et al.*, 2007, Coen *et al.*, 2008, Lindon and Nicholson, 2008). Both NMR and MS methods generate complex datasets that require subsequent multivariate analysis (MVA, as described in section 1.5), to maximize information recovery and possibly biomarker identification (Cloarec *et al.*, 2005, Trygg *et al.*, 2007, Madsen *et al.*, 2009).

1.4 Principles of Nuclear Magnetic Resonance (NMR) spectroscopy and Mass Spectrometry (MS)

1.4.1 NMR Spectroscopy

Proton NMR is based on the magnetic property (or magnetic momentum, μ), originating from their rotating movement, known as spin, and associated to an angular momentum \mathbf{P} and a spin quantum number I . Only the nuclei corresponding to non-zero values of I (which takes integer or half-integer values) are detectable by NMR (Lindon *et al.*, 2000). For instance, ^1H , ^{13}C , ^{15}N , ^9F , ^{31}P have $I=1/2$, ^2H , ^{14}N that have $I=1$ or ^{17}O which has $I=5/2$, are detectable by NMR spectroscopy whilst ^{12}C and ^{16}O that have $I=0$ are not NMR active. For the nuclei with $I \neq 0$, the associated magnetic moment μ is given by $\mu = \gamma \cdot \mathbf{I}$, where γ is the gyromagnetic ratio of the nuclei ($\gamma = 2.675 \times 10^{-8} \text{ rad} \cdot \text{T}^{-1} \cdot \text{s}^{-1}$ for ^1H). For nuclei with $I=1/2$, in the presence of an external magnetic field \mathbf{B}_0 , there are two possible orientations of the spins, corresponding to different energy levels. The number of possible spin orientations is given by the magnetic quantum number m_I and determined according to $m_I = 2I + 1$. These orientations may be either parallel (lower energy, α) or anti-parallel (higher energy, β) (Figure 1.7b) and the energy difference (ΔE) between states α and β is given by $\Delta E = \gamma \cdot \hbar \cdot \mathbf{B}_0$ where \hbar is the Planck's constant, h , divided by 2π . The external magnetic field \mathbf{B}_0 induces a torque (a momentum that causes rotation about an axis) on the nuclei, causing the spins to precess around \mathbf{B}_0 (Figure 1.7a). This motion is known as Larmor precession and it occurs at defined frequency, named the Larmor frequency ν_0 , given by $\nu_0 = -\gamma \cdot \mathbf{B}_0 / 2\pi$ (Hz) or at a rate ω_0 given by $\omega_0 = \gamma \cdot \mathbf{B}_0$ ($\text{rad} \cdot \text{s}^{-1}$).

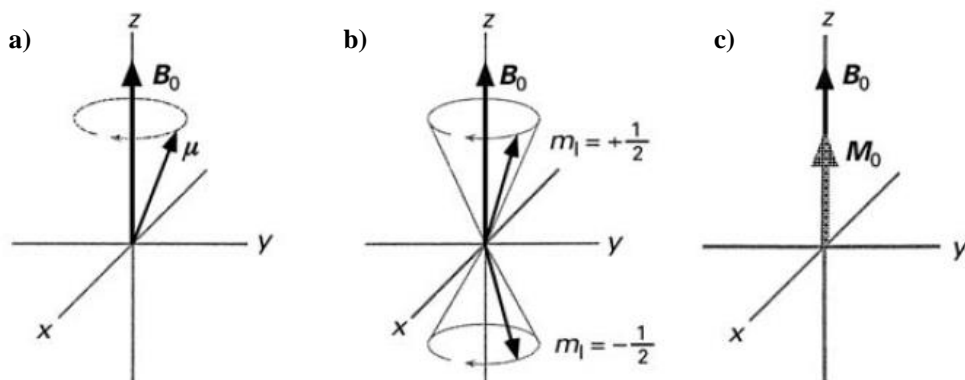


Figure 1.7: a) Precession of an individual magnetic moment μ around the external magnetic field B_0 ; b) precession of magnetic moments in the α ($m_I = 1/2$) and β ($m_I = -1/2$) spin states and c) resultant magnetic moment M_0 reflecting the excess of population in spin state α . Adapted from (Lindon *et al.*, 2000).

The distribution of the spin states by the two energy levels (e.g. in the case of protons) is given by the Boltzmann distribution (Eq. 1.1), where N is the number of spins in each spin state, ΔE is the energy difference between states, k is the Boltzmann constant and T the absolute temperature. According to this distribution, in the presence of B_0 and at thermal equilibrium, there will be an excess of spins in state α thus creating a bulk magnetization, M_0 , parallel to B_0 (Figure 1.7b).

$$\frac{N_\alpha}{N_\beta} = e^{\frac{\Delta E}{kT}} \quad [\text{Eq. 1.1}]$$

Detection of the NMR signal

For an NMR signal to be detected, a radiofrequency (RF) pulse is applied for a given period, usually of the order of μs . This RF has an associated magnetic field, B_1 , oscillating at the same Larmor frequency as the spin. The RF is transmitted via a coil surrounding the sample, which induces a magnetic field B_1 in the transverse plane, thus perpendicular to B_0 (Claridge, 1999). To help visualize the concomitant rotating fields and precessing vectors, a reference rotating frame (x' , y') rotating at the Larmor frequency around z is used. The applied RF pulse (or 90° pulse) causes the net magnetization M_0 to flip from the z axis to the x' - y' plane. After the RF pulse, the magnetization vector is re-established, so that the transverse magnetization gradually disappears towards z returning to the initial value of M_0 , this process is known by relaxation (described below). To detect

the NMR signal, the receiver coil, which is aligned with the $x'-y'$ plane, as show in Figure 1.8a, is crossed by the magnetization vector \mathbf{M}_0 generating a current which is amplified and recorded, known as free induction decay (FID).

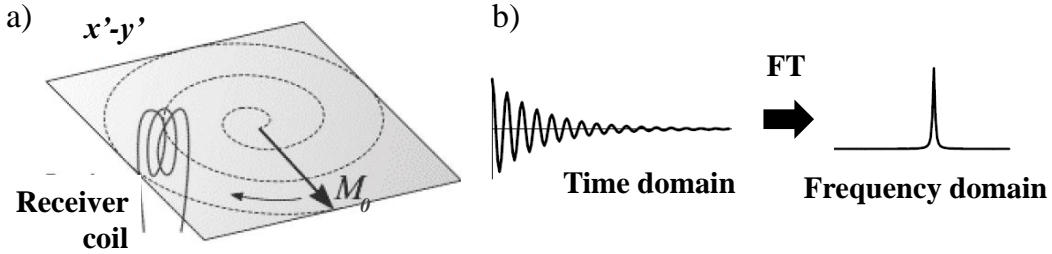


Figure 1.8: a) Scheme of the magnetization vector \mathbf{M}_0 precessing on the $x'-y'$ plane around the z axis, cutting the receiver coil aligned with the $x'-y'$; b) time domain signal of the FID and c) frequency domain signal of the FID after applying FT. Adapted from (Keller, 1988, Keeler, 2002).

The resulting FID is a time domain signal, $f(t)$, which is then transformed into the frequency domain, $f(w)$, i.e. the NMR spectrum, by applying a Fourier transformation (FT, Eq. 1.2).

$$f(w) = \int_{-\infty}^{+\infty} f(t) e^{-iwt} dt \quad [\text{Eq. 1.2}]$$

Relaxation

The relaxation processes can occur according to two main mechanisms, one named as longitudinal (or spin-lattice) and a second process known as transverse (or spin-spin) relaxation, respectively characterized by relaxation times T_1 and T_2 . T_1 and T_2 describe, respectively, the exponential decay of the net magnetization in the z axis and in the $x'-y'$ plane. The longitudinal or spin-lattice relaxation relates to the exchange of magnetization with the surroundings, the z component of the magnetization (M_z) being expressed in terms of T_1 , the time required for recovery, and given by Eq. 1.3.

$$M_z = M_0 (1 - e^{-t/T_1}) \quad [\text{Eq. 1.3}]$$

The transverse or spin-spin relaxation relates to differences in magnetic fields that are experienced by the spins, due to inhomogenities or intra- and inter-molecular interactions in the sample. These differences cause the spanning out of the magnetization component in the $x'-y'$ plane ultimately leading to no net magnetization in this plane and it is expressed in terms of T_2 given by Eq. 1.4 and Eq. 1.5 (Jacobsen, 2007).

$$M_x = M_0 \sin(2\pi w_0 t) e^{-t/T_2} \quad [\text{Eq. 1.4}]$$

$$M_y = -M_0 \cos(2\pi w_0 t) e^{-t/T_2} \quad [\text{Eq. 1.5}]$$

Moreover, T_2 relates to the NMR resonances width, as the half-height line width is proportional to $1/\pi \cdot T_2$ (Claridge, 1999). Thus, large and slowly tumbling molecules will be characterized by large T_2 values and therefore give rise to broad NMR resonances.

Chemical shift and scalar coupling

Each nucleus is surrounded by electrons that create a local magnetic field, opposite to the external magnetic field (\mathbf{B}_0) forming a shielding effect that causes each nucleus to experience a different effective magnetic field (\mathbf{B}_{eff}). The effective magnetic field is given by $\mathbf{B}_{\text{eff}} = \mathbf{B}_0 (1 - \sigma)$, where σ is the shielding constant, which reflects the extent to which the electron cloud shields the nucleus (Lindon *et al.*, 2000, Jacobsen, 2007).

Depending on the chemical environment, e.g. neighboring atoms, chemical bonds, each nuclei within a molecule resonates at different frequencies (ν) resulting in different positions in the NMR spectrum. The resonance frequency of a given nucleus is then divided by the spectrometer frequency (ν_{ref}) and multiplied by 10^6 , resulting in a dimensionless value named the chemical shift (δ) and expressed in parts per million (ppm) (Eq. 1.6). In this way, the resulting chemical shift is constant and irrespective of the frequency of the operating NMR spectrometer. The ppm scale is further calibrated with frequencies of the methyl groups bound to the silicon atom of chemically inert compounds tetramethylsilane (TMS), dimethylsilapentanesulfonic acid (DSS) or trimethylsilyl propionate (TSP), used for ^1H and ^{13}C calibration (Lindon *et al.*, 2000, Jacobsen, 2007).

$$\delta = 10^6 \times \left(\frac{\nu - \nu_{\text{ref}}}{\nu_{\text{ref}}} \right) \quad [\text{Eq. 1.6}]$$

More shielded nuclei, such as those without electronegative atoms (e.g. O, N, Cl) or unsaturated groups (e.g. C=C, C=O, aromatic) in the nearby environment, resonate at lower frequencies, thus situated in the lower ppm scale, whereas resonance of less shielded nuclei, such as those proximal to electronegative or unsaturated groups, are located in the

higher ppm scale (Jacobsen, 2007). An example of some typical ^1H and ^{13}C chemical shift ranges of common groups is shown in Table 1.8.

Table 1.8: ^1H and ^{13}C typical chemical shifts values ranges of common chemical groups. Adapted from (Fan, 1996).

^1H chemical shifts		^{13}C chemical shifts	
Chemical group	δ/ppm	Chemical group	δ/ppm
$\text{C}-\text{CH}_3$	0.7-1.1	$-\text{CH}_3$ (methyl)	8-30
$=\text{C}-\text{CH}_3$	1.5-1.8	$-\text{CH}_2$ (methylene)	14-55
COCH_3 (acetyl)	2.0-2.5	$-\text{CH}-$ (methine)	22-60
$\text{N}-\text{CH}_3$	2.5-3.3	$-\text{C}-$ (quaternary)	30-40
$\text{O}-\text{CH}_3$	3.3-4.3	$\text{CH}_2=\text{C}-\text{R}$	100-150
$\text{CO}-\text{CH}_2-\text{C}$	1.3-2.5	$\text{C}=\text{C}-\text{C}=\text{C}-\text{R}$	110-150
$\text{O}-\text{CH}_2-\text{C}$	4.0-5.0	Heteroaromatic ring	100-165
$\text{C}-\text{CH}-\text{OH}$	3.3-4.0	$\text{C}-\text{OH}$ (alcohol)	44-85
$\text{C}-\text{CH}-\text{O}-\text{ester}$	4.2-5.3	$\text{C}-\text{O}-\text{C}$ (ester)	55-85
$=\text{CH}-\text{C}$ (olefinic)	5.5-8.5	$\text{R}-\text{COOH}$ (saturated)	165-188
Aromatic ring	6.0-9.0	$\text{R}-\text{C}=\text{C}-\text{COOH}$	158-174
$-\text{CHO}$ (aldehyde)	9.0-10.2	$\text{R}-\text{COOR}'$ (saturated)	158-174
$-\text{COOH}$	10.5-13.5	$\text{R}-\text{C}=\text{C}-\text{COOR}'$	152-172
$\text{C}-\text{OH}$ (alcohol)	1.5-6.0	$\text{R}-\text{CHO}$ (saturated)	196-220
$\text{C}-\text{OH}$ (phenol)	6.5-18.5	$\text{R}-\text{C}=\text{C}-\text{CHO}$	176-195
Primary amines	1.1-1.8	Saturated ketones	195-220
Secondary amines	1.2-2.1	$\text{C}-\text{N}-\text{R}_2$	20-70
$\text{CO}-\text{NH}$	5.0-6.5	$\text{R}-\text{CO}-\text{NH}_2$	150-178

Apart from interacting with the external and the radiofrequency magnetic fields, the nuclei also interact with their neighboring nucleus through a process known as spin-spin coupling, scalar coupling or J -coupling which occurs via the electrons in the chemical bonds. This interaction between neighboring spins (e.g. for a given nuclei A and X) causes a shift of their energy levels, depending on their magnetic quantum number m_A and m_X and a parameter that quantifies the strength of the interaction, named coupling constant J_{AX} . The spin-spin coupling causes the NMR signal to split into more than one signals, the multiplicity of which (i.e. number of peaks in the split signal) given by $2nI+1$, where n is the number of neighbor equivalent nuclei (Lindon *et al.*, 2000, Jacobsen, 2007).

Another relevant feature of ^1H NMR spectra is that the area under the peaks are directly proportional to the number of the corresponding nuclei. This means that for a given peak, its integral value is proportional to the number of protons giving rise to that signal. Thus, NMR is an inherently quantitative method as long as the spectra are acquired

in quantitative conditions, that meaning that an inter-scan delay of 5 times the longest T_1 is used, so that complete relaxation of all spins is achieved (Jacobsen, 2007). If the previous quantification condition was taken into account, the concentration of a given compound may be determined following Equation 1.7. This means that the concentration of a given compound, e.g. compound B (C_B), may be calculated when knowing the concentration of a reference compound A (C_A), the integrals of the reference (A_A) and that of the signal of the compound to be quantified (A_B) and the number of protons giving rise to the reference (n_A) and to the signal to be quantified (n_B).

$$\frac{C_A}{C_B} = \frac{A_A \times n_B}{A_B \times n_A} \quad [\text{Eq. 1.7}]$$

In metabolomics studies of complex mixtures, e.g. biofluids, it is difficult to perform absolute quantification due to 1) the broadly overlapped resonances which compromise integration measurements and 2) the limitation of using standard references due to their interactions with sample components (e.g. interactions between TSP and proteins). One alternative is the use of an Electronic Reference To access In vivo Concentrations (ERETIC).

One- and Two-dimensional NMR experiments

One dimensional (1D) NMR experiments usually require a preparation and an acquisition step. The preparation step allows the spins to reach equilibrium (determined by the relaxation delay which corresponds to d_1 parameter and depends on the pulse program), the application of an RF pulse immediately followed by signal acquisition. There are three common 1D ^1H NMR experiments for the study of biofluids or complex mixtures, depending on the composition of such mixtures. Firstly, the “standard” NMR experiment, consisting in a preparation step, with water presaturation, followed by a 90° pulse, is used to detect signals from all molecules whilst other two experiments, based on diffusion coefficients (diffusion-edited) and on T_2 (Carr-Purcell-Meiboom-Gill or CPMG), are used to selectively detect signals arising from macromolecules or from small molecules, respectively (Meiboom and Gill, 1958, Lindon and Nicholson, 2008). The diffusion-edited experiment is performed to detect signals from macromolecules, which are

large and slowly tumbling molecules with high T_2 . In this experiment, an initial 90° pulse is applied followed by the application of gradients, which dephase magnetization. A following gradient delay is introduced to allow molecules to diffuse and, as diffusion depends on the size of the molecule, large molecules will remain in a fairly constant position contrarily to small ones. A subsequent gradient, opposite to the initial is applied, which refocuses magnetization of macromolecules and attenuates signals from smaller molecules. The CPMG experiment, also used to measure T_2 , is performed to selectively detect sharp signals from small molecules, through the use of a spin echo sequence followed by a delay. During this delay period, the magnetization decreases exponentially with the ratio between τ/T_2 . This means that the magnetization from molecules with low T_2 (macromolecules) decreases more rapidly than small molecules with high T_2 . Thus by introducing a τ period before acquisition, broad signals are attenuated.

The 1D spectra (standard and edited) of complex mixtures, such as biofluids, is composed of numerous of peaks, which overlap extensively, hindering peak assignment and spectral interpretation, therefore, two dimensional (2D) experiments are an absolute necessity. 2D NMR experiments are, in general, composed of four steps: preparation, evolution (t_1), mixing and acquisition (t_2). The preparation step is usually a relaxation time to allow spins to reach equilibrium followed by a 90° pulse to bring the magnetization to the $x'-y'$ plane. The second dimension is then established through the evolution step, during which a series of pulses are applied to transfer the magnetization between nuclei. After the evolution time and depending on the experiment, additional pulses may be applied which are then followed by acquisition (Lindon *et al.*, 2000, Jacobsen, 2007).

2D experiments rely on inter-nuclei interactions, which may be homonuclear or heteronuclear, to generate frequency information along two axes and, for metabolomic studies, usually three experiments are performed: Total Correlation Spectroscopy (TOCSY), Heteronuclear Single Quantum Coherence (HSQC), and ^1H NMR J -resolved (J -res), although two other complementary experiments may be also helpful, namely Correlation Spectroscopy (COSY, useful to map homonuclear coupling over two or three bonds) and Heteronuclear Multiple Bond Correlation (HMBC, useful to map heteronuclear coupling across multiple bonds).

While 2D COSY provides coupling information over two or three bonds, the TOCSY spectrum provides further information of coupled protons belonging to the same

spin-system, separated by up to five bonds. Thus, the latter experiment provides a mean of identifying all resonances within a spin system, playing an essential role for metabolite assignment. In the TOCSY experiment, the evolution step is followed by a spin-lock sequence during the mixing time, the MLEV-17 (Malcolm Levitt- 17 sequence) that allows the coherence transfer between all coupled nuclei in a spin-system. The MLEV-17 sequence, consists in 16 composite pulses followed by an uncompensated 180° pulse (Bax and Davis, 1985), the general scheme of the TOCSY sequence being shown in Figure 1.9.

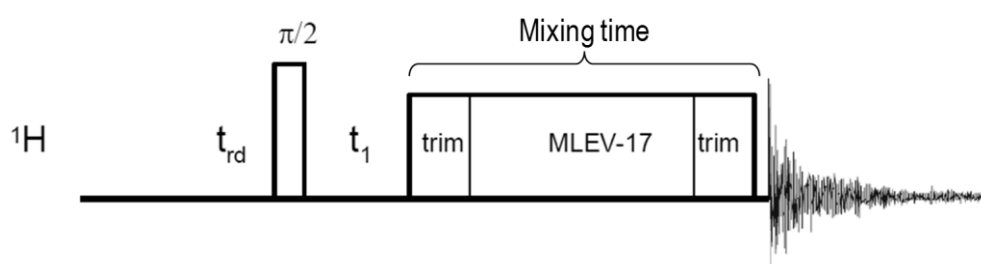


Figure 1.9: TOCSY pulse sequence.

Complementary to TOCSY, the HSQC experiments provide information about the chemical shift of ^{13}C correlated with the ^1H chemical shift of the directly bound proton. The HSQC experiment starts with a 90° pulse for proton magnetization followed by an insensitive nuclei enhanced by polarization transfer (INEPT) sequence that transfers the magnetization from proton to the bounded carbon and also recovers magnetization from carbon to the proton (see Figure 1.10). This sequence is much more sensitive than those involving direct detection of carbon such as the heteronuclear correlation (HETCOR) pulse sequence (Jacobsen, 2007).

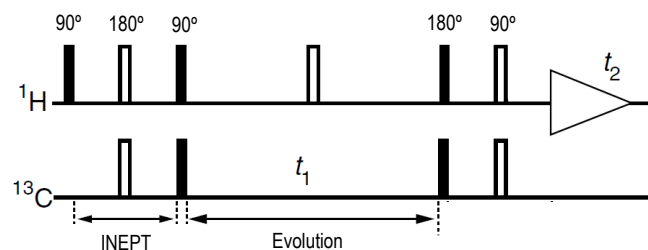


Figure 1.10: HSQC pulse sequence. Figure adapted from (Keeler, 2002).

The J -resolved (or J -res) experiment allows information about chemical shift to be separated from scalar coupling. In this experiment, the 1st dimension (F1) contains

information about multiplicity, whilst scalar couplings are presented in the 2nd dimension (F2), thus enabling the clearer identification of ¹H multiplets without overlapping and allowing the measurement of coupling constants (J). The experiment starts with a 90° pulse to create magnetization and, by the midpoint of the evolution time a 180° pulse is applied (Figure 1.11a) to refocus proton shifts but not scalar couplings, which appear in the 1st dimension (F1). The 2nd dimension (F2) contains information about chemical shift and coupling constants, which complicates the interpretation of the spectra. To retain only chemical shifts in F2 the “tilting” of the multiplets through an angle of 45° is performed, as exemplified in Figure 1.11b (Claridge, 1999). Furthermore, the J -res experiment has been used for quantification studies for complex mixtures.

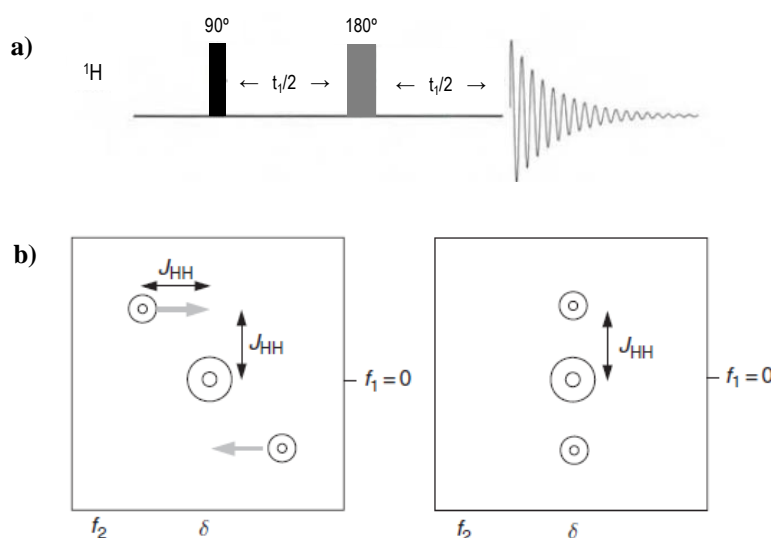


Figure 1.11: J -res pulse sequence. Figure adapted from (Claridge, 1999).

1.4.2 Mass Spectrometry

Mass spectrometry is based on the formation of ionic species through the action of an ionizing energy, these ions being further separated based on their mass to charge ratio (m/z) with intensity proportional to their relative abundance (Hoffmann and Stroobant, 2007). One of the great advantages of MS is its sensitivity and selectivity, as it is able to detect compounds in lower concentration than NMR (up to picomolar to femtomolar). MS also provides specific chemical information such as accurate mass, isotope distribution pattern and characteristic fragment-ions for structural identification and assignment (Lei *et*

al., 2011). Some limitations of MS, however, are its destructive nature and limited use for metabolite quantification as it is prone to variable ionization of different metabolites and ion suppression artifacts (Lindon *et al.*, 2006). Also, metabolite identification using MS can be difficult often requiring the use of tandem MS for ions fragments studies. On the other hand, MS experiments often require a pre-separation step, depending on the type of metabolites on interest, making the technique more advantageous for targeted studies (e.g. detailed study of lipids, etc.). For the analysis of complex mixtures, as is the case biofluids, MS is usually coupled with gas chromatography (GC) to separate volatile compounds, capillary electrophoresis (CE) to separate highly polar and ionic metabolites and liquid chromatography (LC) for the study of polar, less-polar and neutral metabolites (Villas-Bôas *et al.*, 2005, Lei *et al.*, 2011). A recent improvement of the latter separation method is ultraperformance LC (UPLC) which uses higher pressures to achieve higher flow rates, allowing a greater efficiency in metabolites separation to be achieved (Lindon *et al.*, 2004).

Liquid Chromatography-MS (LC-MS)

LC-MS experiments require a high pressure liquid chromatography (HPLC) equipment coupled to an electrospray ionization (ESI) method and MS spectrometer. ESI is produced by applying a potential difference (positive or negative), to a liquid passing through a capillary tube and charged molecules are then sprayed into the gas phase. Because some molecules ionize better in positive and negative modes, both ESI⁺ and ESI⁻ are performed (Hoffmann and Stroobant, 2007). In time-of-flight (TOF) MS equipments, ions in the gas phase are further separated based on their velocity, i.e. after an initial acceleration by an electric field ions enter a field-free region and m/z are determined by measuring the time that ions take to move between the source and the detector (Hoffmann and Stroobant, 2007). In metabolomic studies, usually quadropole time-of-flight (Q-TOF) spectrometers are used due to their stability, sensitivity, rapid acquisition and high mass accuracy (Want *et al.*, 2006, Theodoridis *et al.*, 2012). In the quadrupole detector, ions travel along a path parallel to four voltage carrying rods and, by applying a given RF and voltage current, specific m/z can be selected (Hoffmann and Stroobant, 2007). This equipment is further used for fragmentation experiments, named MS/MS experiments, where ions are selected by a first MS run, further subjected to collision with an inert gas

(Ar or N₂) causing fragmentation, and fragments analyzed by a second MS run, these experiments being essential for structure determination (Hoffmann and Stroobant, 2007).

The separation of the components within a complex biological matrix, as the case of metabolomics, with the traditional HPLC methods is somehow hampered by its lower chromatographic efficiency, sensitivity, resolution and higher analysis time (Want *et al.*, 2010). Thus, a new ultraperformance liquid chromatography (UPLC) is being increasingly used as it utilizes higher pressures and smaller columns with smaller particle size packing material, allowing for improved separation and higher resolution. Moreover, the use of higher pressures results in faster runs, narrower chromatographic peaks, improved signal-to-noise ratio (S/N) and increased sensitivity compared to HPLC, as shown in Figure 1.12 (Want *et al.*, 2006).

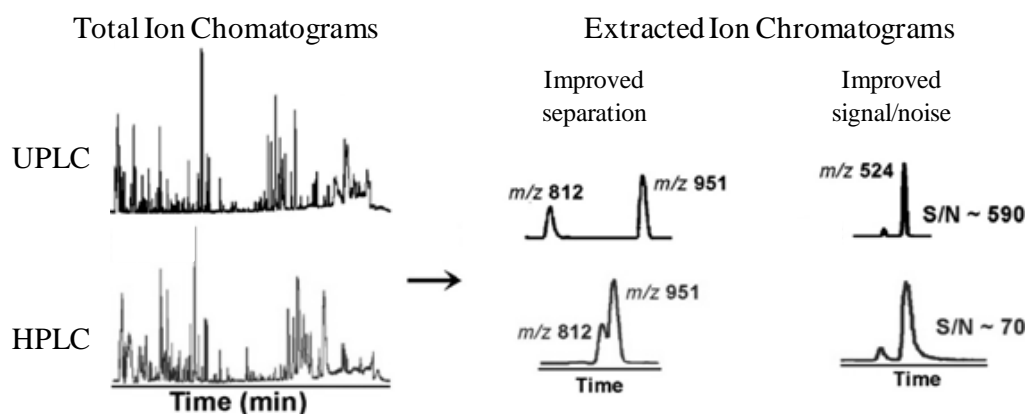


Figure 1.12: Total ion chromatogram and extracted ion chromatogram (EIC) with HPLC and UPLC, showing improved separation of metabolites, improved data interpretability, S/N, accuracy and sensitivity in UPLC (adapted from (Want *et al.*, 2006)).

1.5 Statistical tools in metabolomics

The types of datasets used in metabolomics are inherently large and complex, containing high amounts of potential biological information. The retrieving of relevant and valuable biological information, i.e. subset of variables, and removing data unrelated with the biological condition being studied can be quite a challenge (Saccenti *et al.*, 2013). Typically, metabolomic studies use multivariate and univariate statistical methods, both approaches yielding complementary information. However, prior to the application of the actual statistical tools (multi- and univariate analysis) important preprocessing steps, like alignment, normalization, scaling, transformation, are required to maximize information

recovery and results reliability. Traditionally, both unsupervised and supervised methods are used sometimes with the tandem application of variable selection. The latter methods aim at aiding biological interpretation by removing irrelevant, unreliable or unstructured variations. Moreover, the importance of validating the results has been increasingly recognized and is presently an absolute necessity in metabolomics. Each of these steps will be described in detail throughout this section.

1.5.1 Preprocessing: bucketing, alignment, normalization, scaling and transformation

Bucketing and Alignment

One of the major challenges when dealing with NMR data is the positional variations of peaks in the spectrum, due to changes in pH, temperature, salt concentration or instrumental factors. This becomes even a greater challenge in complex biofluids where different peaks are affected to different extents (Smolinska *et al.*, 2012). For instance, in urine samples there are significant differences in pH and ionic species (Na^+ , K^+ , Ca^{2+} and Mg^{2+}) which causes chemical shifts variations of metabolites with ionizable groups ($-\text{COOH}$, NH_3 , NH_2 , NH). Citrate is an example of a urinary metabolite with high chemical shift variations due to pH. Citrate has three carboxyl groups ($\text{R}-\text{COOH}$) attached to two CH_2 groups, the latter yielding two doublets at 2.54 and 2.69 ppm (chemical structure shown in Figure 1.13). Changes in pH induce differences in $\text{R}-\text{COO}^-$ concentration which

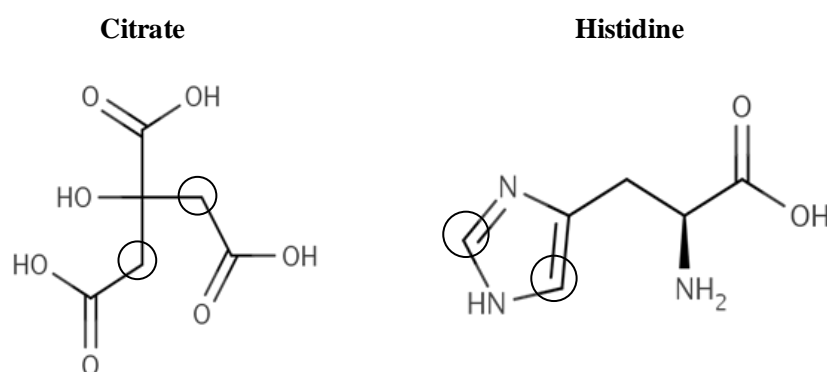


Figure 1.13: Chemical structure of citrate (left) and histidine (right). Circles indicate protons the chemical shift of which is most affected by pH.

alters field-shielding effects on the molecule protons causing shift in the effective field experienced. Resonances of histidine are also highly pH dependent, particularly the two singlets arising from ring protons resonating at 7.13 and 7.98 ppm, due to ionization of the NH group (see chemical structure in Figure 1.13). Chemical shift variation, within the pH range of 7.1-7.7, for the proton resonances of histidine can be up to 0.02 and 0.008 ppm for citrate, while other urinary compounds, namely hippurate, creatinine, dimethylamine, trimethylamine, glycine and acetate, have an associated deviation inferior to 0.002 ppm (Xiao *et al.*, 2009).

To overcome this difficulty, spectral bucketing, or binning, is commonly performed, consisting of integrating the spectrum in fixed or variable widths, with width varying from 0.005 to 0.04 ppm. Spectral bucketing is helpful to eliminate differences in peak position and reducing matrix complexity although at the cost of spectral resolution (Lindon *et al.*, 2007, Smolinska *et al.*, 2012). Alternatively, spectral alignment can be used to solve the problem of peak shifts, while maintaining spectral resolution. Several alignment algorithms are available such as interval correlated shifting (icoshift) (Savorani *et al.*, 2010), recursive segment-wise peak alignment (RSPA) (Veselkov *et al.*, 2009), fast iterative warping algorithm, i.e. fuzzy warping (Alm *et al.*, 2009) and hierarchical cluster-based peak alignment (CluPA) (Vu *et al.*, 2011). An example of spectral alignment is shown in Figure 1.14, highlighting the high chemical shift deviation for citrate resonance (2.5-3.2 ppm), in ^1H NMR spectra of urine, and how it can be corrected using the RSPA algorithm.

After either bucketing or alignment, the spectral dataset are converted into a two dimension data table (or matrix), consisting of n rows corresponding to each observation or sample, and m columns corresponding to each variable, e.g. chemical shift.

Normalization

Normalization is a crucial step in preprocessing datasets for metabolomics as it aims to overcome differences in concentration, dilution or total amount of sample, making all samples within a dataset comparable. This step is particularly important in urine studies where fluctuations in sample concentration can reach several orders of magnitude (Smolinska *et al.*, 2012). Normalization is applied to each row in the matrix (that is each

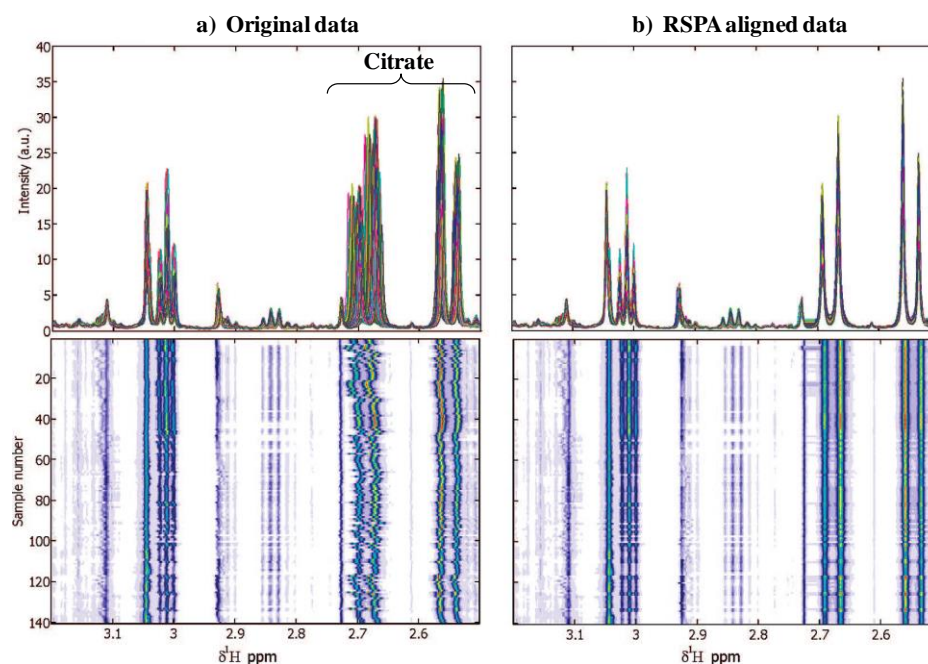


Figure 1.14: Example of peak position variations in the region of 2.5-3.2 ppm of a) original ^1H NMR spectra of urine and b) after RSPA alignment. The lower plots represent the intensity of peak positions in all samples (adapted from (Veselkov *et al.*, 2009)).

spectrum) typically by dividing each row by a constant, there being several ways to compute that constant or normalization factor. A common normalization practice is to divide each variable by the total integrated intensity across the whole profile, known as total sum, total area, total intensity or integral normalization (Veselkov *et al.*, 2011, Smolinska *et al.*, 2012). In urine studies, creatinine normalization is also used as it is based on the assumption that creatinine urinary excretion, i.e. creatinine clearance, is constant (Jankevics *et al.*, 2008, Smolinska *et al.*, 2012). Due to the uncontrolled composition of urine, normalization by total intensity and creatinine normalization poses significant pitfalls as both total sum or creatinine integral can vary irrespectively of sample dilution, for instance if samples contain extreme concentration of endogenous or drug metabolites (Dieterle *et al.*, 2006) or due to creatinine variation caused by changes in diet (Solanky *et al.*, 2005, Wang *et al.*, 2005, Stella *et al.*, 2006), lifestyle (Slupsky *et al.*, 2007), and muscle mass (Craig *et al.*, 2006). To overcome this, other recent normalization methods have been proposed and shown valuable in urine studies, namely probabilistic quotient normalization (PQN) (Dieterle *et al.*, 2006), histogram normalization (HN) (Torgrip *et al.*, 2008), quantile normalization (Veselkov *et al.*, 2011, Kohl *et al.*, 2012), group aggregating

normalization (GAN) (Dong *et al.*, 2011) and locally weighted scatter plot smoothing (LOESS) (Veselkov *et al.*, 2011).

In this work, PQN was the normalization method chosen as it is based on the calculation of the most probable dilution factor, relatively to controls, whilst reducing the contribution of abnormally concentrated metabolites. Briefly, the method comprises the calculation of the quotient between each variable of each spectra and a reference, in this case the median value of the control group. Then, the most probable quotient, i.e. the median value of all quotients, is selected and used as normalization factor for each sample (Dieterle *et al.*, 2006).

Data scaling and transformation

In projection/compression based methods (described in the next section) metabolites present in higher concentrations are given higher importance and can dominate the projections hindering potentially valuable variations in metabolites present in lower concentration (Smolinska *et al.*, 2012, Xia *et al.*, 2013). Thus, scaling is performed to each variable, by dividing the given variable by a scaling factor (Xia *et al.*, 2013). There are several scaling such as Unit Variance scaling (UV, also named auto scaling or standardization), which consists in mean centering (subtracting mean value of the sample) and dividing each variable by its standard deviation, Pareto scaling, consisting in mean centering and dividing each variable by the square root of the standard variation of that variable, and variable stability scaling, or VAST, which is an auto scaling divided by the coefficient of variation of that variable (van den Berg *et al.*, 2006, Smolinska *et al.*, 2012, Xia *et al.*, 2013). Apart from scaling, data transformation, such as log or power transformation, may also be applied to reduce data variability and to make skewed distributions (i.e. asymmetric distributions) more symmetric (van den Berg *et al.*, 2006).

1.5.2 Multivariate analysis methods

In order to achieve efficient and robust modeling methods, to enable analysis of complex biological and biochemical data, and to produce interpretable and reliable models, both unsupervised methods and supervised methods are used (Lindon *et al.*, 2007, Trygg *et al.*, 2007). Unsupervised methods do not require *a priori* knowledge about sample classes,

for instance control vs. disease, maximizing the main sources of variations and allowing the identification of outliers, trends and inherent grouping. Principal Component Analysis (PCA) is undoubtedly the most used unsupervised method in metabolomic studies and will be explained in detail throughout this section. Nonetheless, other unsupervised methods such as hierarchical cluster analysis (HCA), non-linear mapping (NLM), self-organizing maps (SOM) and k -means clustering have been used (Lindon *et al.*, 2007, Smolinska *et al.*, 2012).

Supervised methods, i.e. methods that require *a priori* information about samples membership, maximize systematic differences between the classes and are absolutely essential in metabolomic studies as they enable sample classification and prediction. One of the most widely used supervised method is Partial Least Squares-Discriminant Analysis (PLS-DA), and its orthogonal variations (OPLS-DA) these being explained in detail throughout this section, although other methods, such as Soft Independent Modeling of Class Analogy (SIMCA) and Artificial Neural Networks (ANN) have also been used (Lindon *et al.*, 2007, Smolinska *et al.*, 2012).

Principal Component Analysis (PCA)

PCA is a projection-based method that reduces the multidimensional space into a low-dimensional space, where each projection contains the variation within the \mathbf{X} matrix along each Principal Component (PC). The first PC (PC1) contain the largest variation, the second PC (PC2) explains the second largest variation and so on. All PCs are a linear combination of the original input variables, they are mutually orthogonal and derive from the decomposition of the original matrix \mathbf{X} according to Eq. 1.8, where \mathbf{T} is the matrix of projections of \mathbf{X} , that is the scores, \mathbf{P} is the loadings matrix and \mathbf{E} is the residual matrix.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad [\text{Eq. 1.8}]$$

The visualization of sample similarities explained by the PC is carried out with basis on the scores and loadings plots. The scores plot represents the observations, the samples, with the position of each observation being given by the relative similarity between samples, hence similar observations are positioned closer in the scores space. In this way, a scatter plot of the first 2 PC's provides a summary of all observations in the dataset, allowing the identification of grouping, trends or outliers. The loadings plot shows

the relation between variables, providing a mean to visualize the individual contribution of each variable, e.g. chemical shifts or m/z , in the model (Lindon *et al.*, 2007, Trygg *et al.*, 2007, Smolinska *et al.*, 2012). A general scheme of the construction of a PCA model from a multidimensional dataset is shown in Figure 1.15. Here, the multidimensional dataset (\mathbf{X} matrix) is projected into a two-dimensional plane, with the two first components containing the main variation within the \mathbf{X} matrix. Relevant information is then recovered by means of the scores plot, which shows those observations that are similar/different, and of the loadings plot, showing which variables contribute the most for the similarity/differences between the observations.

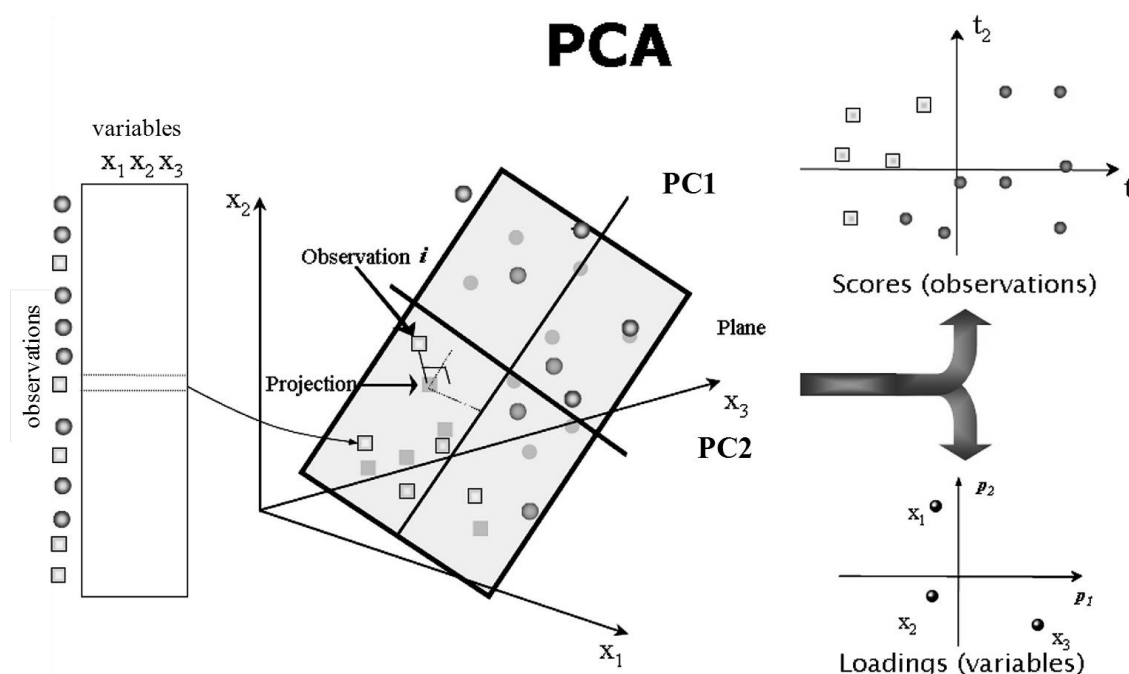


Figure 1.15: Scheme of the construction of a projection-based PCA model (adapted from (Trygg *et al.*, 2007)).

Partial Least Squares Discriminant Analysis (PLS-DA) and Orthogonal PLS-DA (OPLS-DA)

The PLS method measures the quantitative relationship between two matrices, namely between the \mathbf{X} matrix, which contains analytical data, and the \mathbf{Y} matrix containing quantitative values, e.g. measurements of an endogenous metabolite. The method can also be used to discriminate between sample classes, named PLS-DA, and in that case the \mathbf{Y} matrix contains information about sample membership, i.e. controls *vs.* disease (Lindon *et al.*, 2007). PLS-DA, similarly to PCA, is a projection-based method that assumes that data

can be approximated in a low-dimensional space by decomposing variation into latent variables (LV), which are linear combinations of the original variables. The first LV in a PLS-DA model contains the highest source of variation between classes, which is not necessarily the highest source of variation within the dataset. (Lindon *et al.*, 2007, Trygg *et al.*, 2007, Smolinska *et al.*, 2012). The model can be expressed in terms of \mathbf{X} and \mathbf{Y} as given by Eq. 1.8 and Eq. 1.9, where \mathbf{T} represent the scores common to the \mathbf{X} and \mathbf{Y} matrix respectively, \mathbf{P} and \mathbf{C} represent the loadings of \mathbf{X} and \mathbf{Y} matrix respectively and \mathbf{E} and \mathbf{F} the respective residual information in \mathbf{X} and \mathbf{Y} (Smolinska *et al.*, 2012).

$$\mathbf{Y} = \mathbf{T}\mathbf{C}^T + \mathbf{F} \quad [\text{Eq. 1.9}]$$

Furthermore, the \mathbf{T} scores matrix is a linear combination of the \mathbf{X} matrix and the coefficients weights \mathbf{W} , and can be given by Eq. 1.10.

$$\mathbf{T} = \mathbf{X}\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1} \quad [\text{Eq. 1.10}]$$

Thus, if combining Eq. 1.9 and Eq. 1.10 the \mathbf{Y} matrix can be expressed in terms of \mathbf{X} (Eq. 1.11) named regression equation, where b represents the regression coefficients of each variable (Wold *et al.*, 2001).

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{F} \quad [\text{Eq. 1.11}]$$

Orthogonal-PLS-DA or OPLS-DA is a modification of the PLS-DA method which has also been used in metabolomics. In this case, the systematic variation between the \mathbf{X} matrix is split into two parts, one is linearly related to \mathbf{Y} , and thus to sample classes, and one that is unrelated to \mathbf{Y} (named orthogonal) which is removed. OPLS-DA is a useful tool for model interpretation (Smolinska *et al.*, 2012) although for prediction purposes PLS-DA should be used (Tapp and Kemsley, 2009).

Data correlation tools: Statistical Total Correlation Spectroscopy (STOCSY) and Statistical Heterospectroscopy (SHY)

Statistical Total Correlation Spectroscopy, or STOCSY, is a useful tool for peak assignment and measurement of metabolite correlations. STOCSY is based on the multicollinearity of peaks intensities in a set of spectra, to generate a pseudo-two-dimensional map that shows the correlation among spectrum intensities, allowing the

identification of positive correlations between peaks that arise from the same molecule. STOCSY may also provide a mean to identify molecules that interact in the same metabolic pathways, because of their biological covariance, by showing a positive or negative inter-molecular correlation. The method is computed based on the properties of the correlation matrix \mathbf{C} , given by Eq. 1.12, where \mathbf{X}_1 and \mathbf{X}_2 denote the autoscaled data matrices (Cloarec *et al.*, 2005).

$$\mathbf{C} = \frac{1}{n-1} \mathbf{X}_1^T \mathbf{X}_2 \quad [\text{Eq 1.12}]$$

STOCSY of NMR data was firstly introduced to measure correlations between 1D ^1H NMR spectra, the methodology proving its value for metabolite structure assignment and interpretation of metabolic variations (Cloarec *et al.*, 2005). An example of the application of STOCSY for metabolite assignment is shown in Figure 1.16a) where an NMR spectrum of urine is colored accordingly to the correlation coefficient of the resonance at 2.512 ppm with the whole spectral region. This means that highly correlated resonances, approximately at 2.90, 7.20 and 7.40 ppm (colored in red), arise from the same metabolite.

The concept of STOCSY has further evolved and extended to the study of correlation between different nuclei, namely ^1H and ^{13}P named heteronuclear or HET-STOCSY (Coen *et al.*, 2007, Wang *et al.*, 2008) and correlation between NMR experiments, e.g. diffusion-edited *vs.* CPMG (Wang *et al.*, 2008). An example of HET-STOCSY is shown in Figure 1.16b, where a 2D map of the correlation between ^{31}P and ^1H of CPMG magic angle spinning NMR spectra of intact intestinal tissue is depicted. In this plot, ^1H and ^{31}P resonances with high correlation belong from the same molecules.

Statistical correlation may also be applied for the measurement of metabolic relationships between data obtained by different analytical platforms. The method is named Statistical Heterospectroscopy or SHY and it has enabled the correlation of NMR *vs.* MS (Crockford *et al.*, 2006, Graça *et al.*, 2012) and NMR *vs.* mid-infrared spectroscopy (MIR) data of the same biofluid (Graça *et al.*, 2013), to correlate NMR data of different biofluids (plasma *vs.* cerebrospinal fluid-CSF) and biofluids NMR spectra with magnetic resonance spectroscopy (MRS) measurements *in vivo* of brain tissue (Maher *et al.*, 2011). An example of inter-techniques correlation is shown in Figure 1.16c) where the correlation

between urine ^1H NMR spectra and the corresponding MS data, that is the same samples analyzed by different techniques, is shown. In this case, high correlation between a given m/z and NMR peaks suggest that they belong to the same compound. Another example of SHY is shown in Figure 1.16d), this figure showing the correlation between two biofluids (plasma and CSF). This plot enabled the identification of resonances from the same peaks detected in different matrices, as exemplified for propylene glycol, valine and lipoproteins.

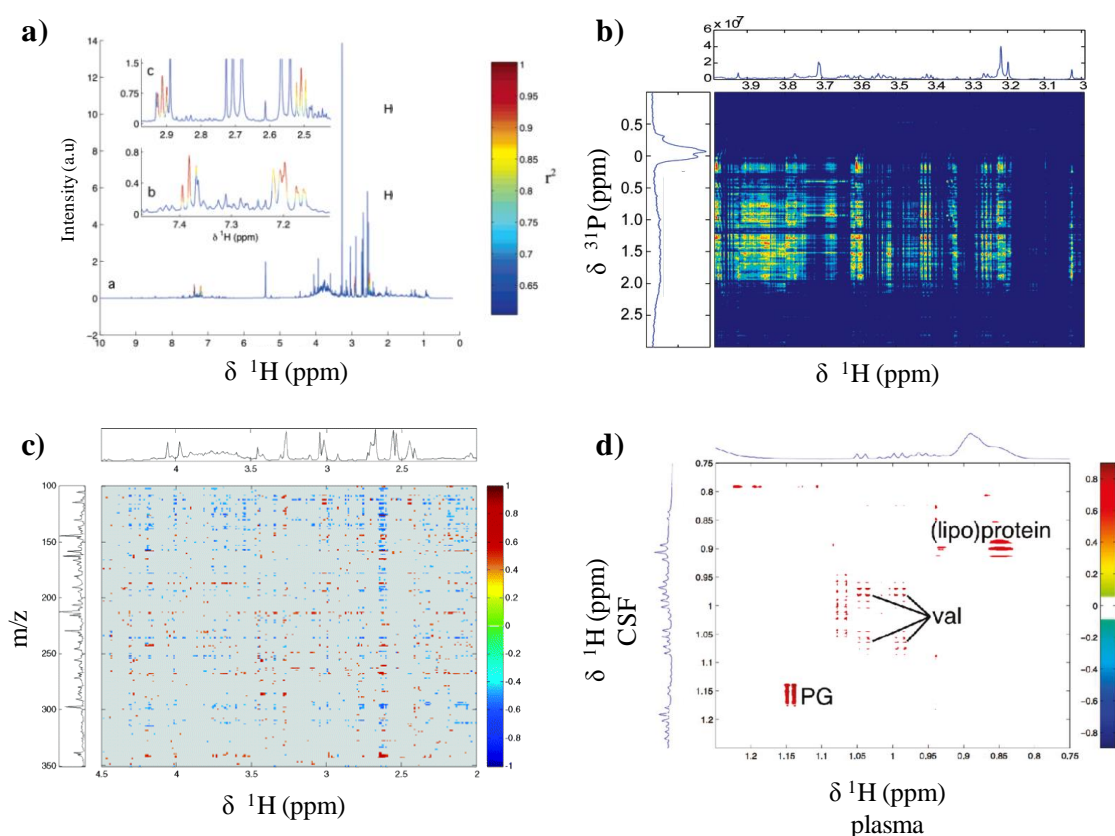


Figure 1.16: a) 1D STOCYSY of the variable at $\delta 2.512$ ppm correlated with the whole spectrum of urine colored by the correlation coefficient; b) 2D heteronuclear STOCYSY (HET-STOCYSY) of ^1H vs. ^{31}P CPMG magic angle spinning NMR of intact intestinal tissue; c) SHY map of ^1H NMR vs. MS data (m/z 100-350) of urine; d) SHY map of CPMG ^1H NMR spectra of plasma vs. cerebrospinal fluid (CSF). PG: propylene glycol, val: valine. Adapted from (Cloarec *et al.*, 2005, Crockford *et al.*, 2006, Wang *et al.*, 2008, Maher *et al.*, 2011).

1.5.3 Univariate statistical methods

Univariate statistical methods are mostly insufficient for metabolomic studies as these only provide comparison for individual variables without taking into account the

relationship between them. Still, they are commonly applied and often used to support multivariate (MVA) results. It is important to apply both uni- and multivariate statistical methods, as they may provide complementary information, but not to confirm each other. This is because MVA provides information about metabolic relationships which are mirrored through metabolites correlation and covariation. On the other hand, univariate analysis only compares individual variations without taking into account the rest of the profile. Thus, when variable intensities are compared individually the relationship between metabolites is not considered and the result can differ from that obtained through MVA. This means that a given metabolite can be found relevant through MVA and not be individually statistically significant when the rest of the profile is not considered. The opposite holds also true, that is individual metabolites can be found to be significantly different between groups while no relevant changes are unveiled through MVA. A possible explanation resides on the complexity of the dataset which can contain too many uninformative variables that mask relevant information (Saccenti *et al.*, 2013).

Univariate methods are based on hypothesis testing, meaning that they comprise a) the definition of the null hypothesis i.e. no difference between the populations' means or medians and b) the probability threshold of the null hypothesis (p -value) rejection (Vinaixa *et al.*, 2012). These methods can be parametric or non-parametric if the data follows a normal distribution or not, respectively. Thus, the choice of the method is strongly dependent on the type of groups and distribution of the data. A compilation of the most appropriate test for different data distributions and group definition is shown in Table 1.9. Normally distributed data is usually compared through Student t-test (or just t-test) to compare two groups, or ANOVA (acronym for analysis of variance), to compare more

Table 1.9: Summary of univariate statistical tests for datasets following normal and non-normal distributions. ^a Paired groups mean that samples were collected from the same subjects for different conditions. Adapted from (Petrie and Sabin, 2009, Vinaixa *et al.*, 2012).

	Normally distributed data (parametric test)	Non-normally distributed data (non-parametric test)
2 paired ^a groups	Paired t-test	Wilcoxon signed rank test
2 independent groups	Unpaired t-test	Wilcoxon rank sum test
> 2 paired ^a groups	Repeated-measures ANOVA	Mann-Whitney Friedman
> 2 independent groups	One-way ANOVA	Kruskal-Wallis

than two groups. Conversely, non-parametric tests are used to compare data that do not follow a normal distribution. The comparison of non-normally distributed data is usually done by applying Wilcoxon signed rank test, Wilcoxon rank sum test, Mann-Whitney test, Kruskal-Wallis, Friedman test can be used (Petrie and Sabin, 2009, Vinaixa *et al.*, 2012).

All aforementioned tests provide a p -value, which reflects the probability of wrongly rejecting the null hypothesis. Usually p -value <0.05 , or significance level of 95%, is used, meaning there is sufficient evidence to reject the null hypothesis in 95% of cases (whilst the opposite, $p>0.05$ means that there is no sufficient evidence to reject the hypothesis) (Petrie and Sabin, 2009).

1.5.4 Variable selection methods

One of the exquisite advantages of NMR- and MS- based metabolomics is the simultaneous detection of hundreds of metabolites in one single run. However, when dealing with complex samples, such as biofluids, the NMR and MS matrices obtained are particularly challenging to work with due to the vast amount of information they contain. In such cases, careful simplification of the original dataset by removing irrelevant, unreliable or noisy variables through variable selection methods may lead to improved model performance, reduced model complexity and biochemical interpretation (Andersen and Bro, 2010).

There are several methods to perform variable selection, the simplest way, but still uncommonly used, being the selection of variables based on an univariate test (Smolinska *et al.*, 2012). The limited applicability of these methods is due to univariate nature of the tests which imply the lost of relationships between variables, thus possibly compromising the multivariate profile insight. In most studies, the selection is based on multivariate model parameters, namely loadings, b -coefficient and VIP values, but also interval PLS (iPLS) (Andersen and Bro, 2010, Smolinska *et al.*, 2012). Selection based on loadings, b -coefficient and VIP values usually encompasses the definition of a threshold value that defines if a given variable is selected or removed. iPLS involves the division of the spectra in intervals of equal length and the computation of PLS models for each of those intervals in order to select the region or set of regions that give the best result. Out of the abovementioned PLS-DA model parameters, selection based on a VIP criteria is the mostly used. The mathematical definition of the VIP value (given by Eq.1.13) provides a

combination of how well a given variable describes the \mathbf{X} (i.e. metabolites) and \mathbf{Y} matrix (sample class). In that equation w_{jf}^2 is the weight value of variable j in component f , SSY_f is the sum of squares of explained variance in component f , J is the total number of variables, SSY_{total} is the total sum of squares explained by the dependent variable \mathbf{Y} and F is the total number of components. A VIP value smaller than 1 means that the variable is not important for the projection so that it can be removed and vice-versa (Andersen and Bro, 2010, Quintás *et al.*, 2012, Sun *et al.*, 2012).

$$VIP_j = \sqrt{\frac{\sum_{f=1}^F w_{jf}^2 \cdot SSY_f \cdot J}{SSY_{total} \cdot F}} \quad [\text{Eq. 1.13}]$$

In metabolomic studies, variable selection methods have been applied to reduce datasets obtained for different biological matrices (e.g. urine, blood plasma and serum, liver tissue) in disease research (Cavill *et al.*, 2009, Lin *et al.*, 2011, Quintás *et al.*, 2012, Sun *et al.*, 2012, Kaur *et al.*, 2013) but also for food quality control (Di Anibal *et al.*, 2011). A variant of this strategy is based on the fact that different selection methods yield different sets of selected variables and, if the commonly selected variables are considered i.e. the intersection of variables selected by different methods, the consistently relevant information may be retrieved and analyzed as a potential profile of the condition under study. This principle has been applied to the study of non-alcoholic fatty liver disease, where variables selected using both regression coefficients and VIP were considered as potential biomarkers of the disease (Quintás *et al.*, 2012). A slightly different concept is the use of combination of selected features instead of intersection. This approach has been applied to the study of liver diseases, where variables were selected using Support Vector Machines-recursive feature elimination (SVM-RFE), genetic algorithm and random forest (RF) (Lin *et al.*, 2011) and to study type 2 diabetes mellitus (T2DM) by selecting variables based on their loading correlation ($p[\text{corr}]$) and RF (Kaur *et al.*, 2013).

1.5.5 Validation methods

In spite of apparent geometrical sample groups separation obtained through PLS-DA and OPLS-DA models' scores plots, these can be meaningless to evaluate differences

between groups (Westerhuis *et al.*, 2008). Thus, determining model robustness and predictive ability, through validation, is an indispensable step in metabolomic studies. A typical study involves two classes, for instance control and disease and, ideally, samples should be split into a training set, used to develop and test the model, and an independent set, used to verify if the model accurately classifies a new sample in the correct group, e.g. control or disease. This type of validation, named external validation, is the ultimate and best manner to validate a multivariate model (Westerhuis *et al.*, 2008, Smolinska *et al.*, 2012).

When dealing with human health, sample numbers are commonly low and the whole sample groups needs to be used as the training set. To overcome this limitation there are other ways to validate models without the use of an independent group. For instance, internal Cross Validation (CV) indicates how well a model works in predicting new samples thus the model's prediction power. CV is a reliable way to test model predictive ability and consists in splitting the data into a number of groups (G), i.e. five to nine, and developing parallel training models with all except one group ($G-1$) leaving that group to be used as the prediction test (Wold *et al.*, 2001, Westerhuis *et al.*, 2008). Another way to validate the model is to perform permutation tests, which indicates if a given model is significantly different from a null hypothesis model, obtained by random guessing, for a given population, thus the robustness of the model (Xia *et al.*, 2013). Permutation tests imply that samples class labels, i.e. control and disease, are randomly permuted and models are recomputed. If the original model is robust, the newly calculated model, after samples class permutation, should not be able to correctly predict classes (Westerhuis *et al.*, 2008).

There are other quality assessment parameters to evaluate models performance and predictive ability. The default parameter to evaluate PLS-DA discrimination ability is the predictive power Q^2 , which is computed according to Eq. 1.14, where \hat{y}_i refers to the predicted class membership for sample i and \bar{y} to the mean y value for all samples. Ideally, the value of Q^2 should equal 1, meaning that predicted classes always matched original classes. However due to the inherent variation within- and between classes it is difficult to achieve the ideal Q^2 value. Hence, a good manner to evaluate the predictive power of the model is by evaluating the Q^2 distribution of permuted models and comparing the result with the original Q^2 . Minimal overlapping of Q^2 values of true and permuted models

indicates a model with predictive power (Westerhuis *et al.*, 2008) as exemplified in Figure 1.17a.

$$Q^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad [\text{Eq 1.14}]$$

Other helpful parameters to evaluate model robustness comprise the classification rate, sensitivity and specificity, given in percentage values, and computed by means of confusion matrices. A confusion matrix consists in the number of false positives (FP), false negative (FN), true positives (TP) and true negatives (TN) as shown in Figure 1.17b. The classification rate refers to the percentage of samples correctly classified in the model, as given by Eq. 1.15. The *sensitivity*, or *true positive rate (TPR)*, refers to the number of the true positives as a percentage of all positives, calculated by Eq. 1.16. *1-specificity* or *false positive rate (FPR)*, refers to the number of false positives as a percentage of all negatives, calculated according to Eq 1.17.

$$CR = \frac{TP + TN}{TP + TN + FP + FN} \quad [\text{Eq. 1.15}]$$

$$\text{Sensitivity (TPR)} = \frac{TP}{TP + FN} \quad [\text{Eq. 1.16}]$$

$$1\text{-specificity (FPR)} = \frac{FP}{TN + FN} \quad [\text{Eq. 1.17}]$$

The plotting of *sensitivity (TPR)* vs. *1-specificity (FPR)* leads to the receiver operating characteristic (ROC) plot. The traditional ROC plot (shown in Figure 1.17c) provides a mean of a) determining if a given test provides useful information on class membership (given by sensitivity and specificity) and b) selecting the optimal cut-off value for a that given test. To obtain a ROC curve, each pair of *sensitivity* and *1-specificity* are obtained for all cut-off values of the given test and connected forming a curve (an example being shown in Figure 1.17c). The obtained curve may be used to determine the accuracy of that test by considering the area under the curve (AUC). The greater the AUC expresses the higher test accuracy for discriminating between classes. This means that an AUC of 1 (ideal value), reflects perfect discrimination between classes whilst an AUC 0.5 reflects a non-discriminating test (Fawcett, 2006, Petrie and Sabin, 2009). In this work, however, a slightly different ROC was used (shown in Figure 1.17c), as they were computed considering the sensitivity and specificity of each PLS-DA model, i.e. considering the whole profile and calculated through Monte Carlo Cross Validation (MCCV). MCCV

consists in splitting the data into a training and a prediction set, the latter set being used for class membership prediction. One advantage of using the MCCV method is that it usually performs 500 to 1000 iterations and calculates for each iteration the Q^2 value and confusion matrix (sensitivity and specificity). The MCCV can be further repeated for permuted classes, i.e. the \mathbf{Y} matrix is randomly permuted, and if the model is robust a complete separation of true and permuted classes is seen in the ROC plot (Xu and Liang, 2001, Xia *et al.*, 2013).

An example of a ROC plot of an ideal model, obtained by MCCV, is shown in Figure 1.17d where it is visible that models obtained with true classes have sensitivity and specificity equal to 1 (100%) and are clearly separated from the permuted classes, which have both sensitivity and specificity 50% (Westerhuis *et al.*, 2008).

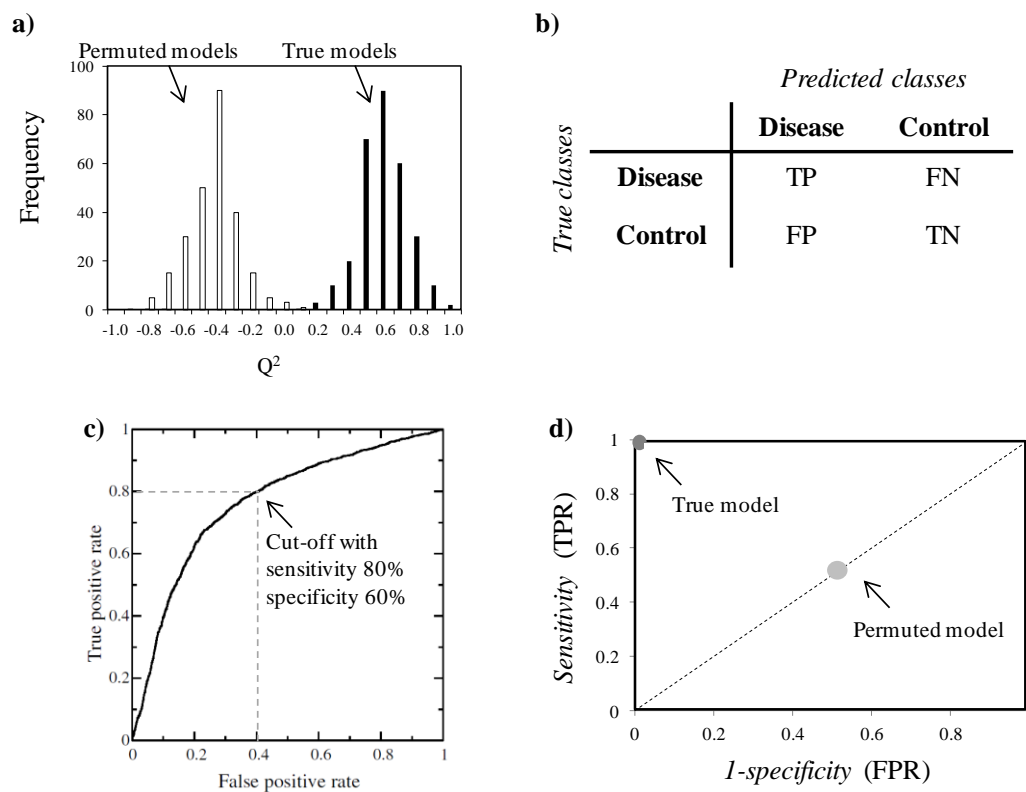


Figure 1.17: Examples of a) Q^2 distribution of true and permuted models, b) confusion matrix obtained for two classes. TP: true positive, FN, false negative, FP: false positive, TN: true negative and c) traditional and d) MCCV-obtained representations of ROC space plot, TPR: true positive rate, FPR: false positive rate.

1.6 State of the art of metabolomics in prenatal research

The merit of metabolomics applied to prenatal health research has been recognized, as reflected through the number of publications in this field. Figure 1.18 illustrates the number of papers published in relation to metabolomics (based on NMR, MS, HPLC and MIR) of pregnancy disorders, these comprising maternal, fetal and newborns conditions, plotted as a function of a) publication year, b) biological matrix and c) disorder studied. An increasing number of publications is clearly visible from 2011 until the present year, whilst until 2010 (coinciding with the beginning of this thesis) a scarce number of papers had been reported in literature (Figure 1.18a), supporting the pertinence of this work. Regarding the biological matrices used in these studies, maternal blood (plasma and serum) has been the most explored biofluid, followed by amniotic fluid, newborn umbilical cord blood, maternal urine (four of which resulted from the work presented in this thesis), placenta and newborn dried blood spots (Figure 1.18b) in decreasing order. The mostly studied disease was found to be PE, followed by FM, PTD, GDM and maternal hyperglycemia, CD and perturbations of intrauterine growth (including SGA, IUGR, Low birthweight-LBW and very low birthweight-VLBW) as shown in Figure 1.18c. A brief review of the main results reported for each disorder, is presented below whilst a comprehensive list of papers (including number of samples, gestational age at sampling, analytical technique used, statistical approach and main biochemical findings) is presented in Annex I, Table A-I.1 (page 293), grouped by disorder and in decreasing order of number of reports per disorder.

Preeclampsia

Previous studies have reported the effect of PE, mostly at post-diagnosis state, on the metabolic profile of AF (Bock, 1994), placenta extracts (Jain *et al.*, 2004, Dunn *et al.*, 2009, Dunn *et al.*, 2012) and maternal blood serum and plasma (Kenny *et al.*, 2005, Turner *et al.*, 2007, Kenny *et al.*, 2008, Turner *et al.*, 2008, Turner *et al.*, 2009). Main findings include hypoxia-related changes, oxidative stress and lipids perturbations (Jain *et al.*, 2004, Turner *et al.*, 2007, Kenny *et al.*, 2008, Turner *et al.*, 2008, Dunn *et al.*, 2009, Dunn *et al.*, 2012). Apart from the characterization of the biochemical phenomena accompanying PE, an increasing effort towards unveiling early and potential metabolic markers of PE has

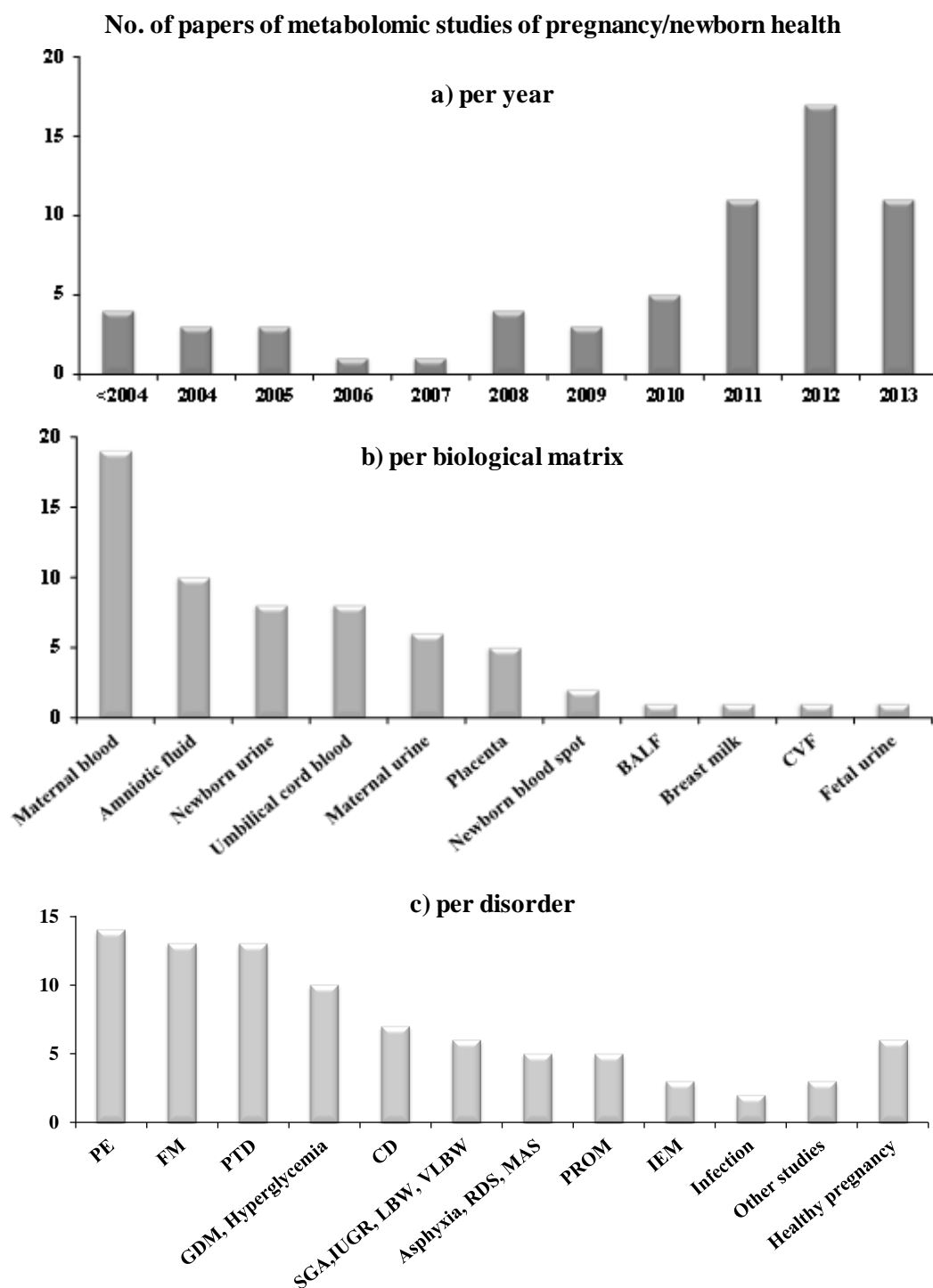


Figure 1.18: Number of NMR, MS, MIR and HPLC metabolomic studies in relation to pregnancy disorders, including maternal, fetal and newborn disorders, shown by a) year published, b) biological matrix used and c) disease studied. BALF: bronchoalveolar lavage fluid, CVF: cervico-vaginal fluid, PE: preeclampsia, FM: fetal malformations, PTD: preterm delivery, PROM: premature rupture of the membranes, CD: chromosomal disorders, GDM: gestational diabetes mellitus, SGA: small for gestational age, IUGR: intrauterine growth restriction, LBW: low birth weight, VLBW: very low birth weight, RDS: respiratory distress syndrome, MAS: meconium aspiration syndrome, IEM: inborn error of metabolism.

been found in several recent publications, using maternal blood serum and/or plasma (Kenny *et al.*, 2010, Odibo *et al.*, 2011, Bahado-Singh *et al.*, 2012, Bahado-Singh *et al.*, 2012) and maternal urine, in a publication resultant of this work (Diaz *et al.*, 2013), collected several weeks prior to diagnosis. It has been suggested that a 14-metabolite signature in 2nd trimester maternal plasma discriminates women further developing PE with an AUC of 0.92. This signature comprises amino acids, carbohydrates, carnitines, dicarboxylic acids, fatty acids, ketones, keto- and hydroxyl- fatty acids, lipids, phospholipids and steroids (Kenny *et al.*, 2010). A subsequent report found that hydroxyhexanoylcarnitine, alanine, phenylalanine and glutamate were significantly higher in 1st trimester maternal plasma of women later developing PE (Odibo *et al.*, 2011). Furthermore, the combination of the latter 4 metabolites yielded an AUC of 0.82 for prediction of PE and 0.85 for early-onset PE (Odibo *et al.*, 2011). More recently, a couple of studies reported significant changes in 1st trimester maternal blood serum in relation to early- and late- onset PE (Bahado-Singh *et al.*, 2012, Bahado-Singh *et al.*, 2012). The authors reported that early-onset could be predicted based on citrate, glycerol, hydroxyisovalerate and methionine (Bahado-Singh *et al.*, 2012) whilst late-onset PE was related with changes in valine, pyruvate, 3-hydroxybutyrate, 1-methyl-histidine, glycerol and trimethylamine (Bahado-Singh *et al.*, 2012). Moreover, it was found that early- and late-onset PE were distinguishable based on glycerol, acetate, trimethylamine and succinate at the 1st trimester (Bahado-Singh *et al.*, 2012).

Fetal malformations

A significant impact on amniotic fluid (AF) composition has been reported in FM cases, considering heterogeneous groups (Graça *et al.*, 2009, Graça *et al.*, 2010, Graça *et al.*, 2012, Graça *et al.*, 2012, Graça *et al.*, 2013) and specifically for spina bifida (Bock, 1994, Groenen *et al.*, 2004). AF studies of FM has unveiled derangements in the energy metabolism under hypoxia, with higher gluconeogenesis and ketogenesis, deficient use of the respiratory chain, changes in polyol and lipids metabolism, the latter with a further involvement in membrane phospholipids acting as lung surfactants. Furthermore, a significant perturbation in amino acids metabolism have also been found, specifically the methionine metabolism in connection with the folic acid pool, and glutamine/glutamate

ratio with connection to fetal kidney underdevelopment (Groenen *et al.*, 2004, Graça *et al.*, 2009, Graça *et al.*, 2010, Graça *et al.*, 2012, Graça *et al.*, 2012, Graça *et al.*, 2013).

A more systemic metabolic effect of FM has also been evaluated through the analysis of maternal blood serum and plasma of women carrying fetuses with malformations (Diaz and Pinto *et al.*, 2011) and specifically for cases of neural tube defects (NTD) (Zheng *et al.*, 2011, Liang *et al.*, 2012). Changes in energy production and lipids conversion along with a derangement in methionine metabolism (in connection with the folic acid pool as suggested in the previous AF studies mentioned above), impaired mitochondrial respiration and changes in the neurotransmitter γ -aminobutyrate (or GABA), neurogenesis and oxidative stress has also been found (Diaz and Pinto *et al.*, 2011, Zheng *et al.*, 2011, Liang *et al.*, 2012). Regarding the use of maternal urine for the study of FM, three studies of samples collected at 2nd trimester (Diaz and Pinto *et al.*, 2011, Graça *et al.*, 2012, Diaz *et al.*, 2013) have been reported, two resulting from the work here presented. Changes were consistent with previously advanced hypothesis of enhanced gluconeogenesis under hypoxia and TCA cycle demand along derangements in methionine metabolism, whilst a further perturbation in nucleotide metabolism was newly put forward (Diaz and Pinto *et al.*, 2011, Graça *et al.*, 2012, Diaz *et al.*, 2013).

Preterm delivery (PTD)

The possibility of predicting the onset of PTD has also been subject to several researches. In this context, the quest for potentially predictive markers of the condition has been pursued through the analysis of AF collected at the 2nd trimester (Graça *et al.*, 2010, Romero *et al.*, 2010, Graça *et al.*, 2012, Graça *et al.*, 2013), as well as in a preliminary study of cervico-vaginal secretions (CVS) (Auray-Blais *et al.*, 2011) and maternal urine (Diaz and Pinto *et al.*, 2011, Diaz *et al.*, 2013), the latter resulting from the work presented in this thesis. AF studies has unveiled changes in amino acids, glucose, ketone bodies, allantoin, choline, proteins and glycoproteins, suggestive of perturbations in amino acid fluxes in the feto-placental unit, increased oxidative stress, hyperglycemia, methionine metabolism and changes in glycoprotein content (Graça *et al.*, 2010, Romero *et al.*, 2010, Graça *et al.*, 2012, Graça *et al.*, 2013). Studies on umbilical cord and maternal blood composition in cases of PTD of VLBW newborns revealed low levels of acetate, and increased lipids, pyruvate, glutamine, valine and threonine in maternal plasma, whilst

decreased lipoproteins, pyruvate and albumin-lysyl and increased glutamine were found in umbilical cord blood reflecting an altered maternal/fetal/placental symbiosis (Tea *et al.*, 2012). Still, the previous findings reflect the concomitant effect of gestational age and birth weight as cases were preterm VLBW. Regarding the newborn, being born preterm has been explored through the analysis of neonatal urine collected in the first days of life (Foxall *et al.*, 1995, Trump *et al.*, 2006, Atzori *et al.*, 2011). Several changes have been found, suggestive of altered renal function, changes in tryptophan-NAD⁺ pathway and tyrosine, arginine and proline metabolism, urea cycle and tyrosine, tryptophan and phenylalanine biosynthesis (Foxall *et al.*, 1995, Trump *et al.*, 2006, Atzori *et al.*, 2011). In spite of these advanced hypotheses, the knowledge on the biochemical impact of being born preterm is still very limited.

Gestational diabetes mellitus (GDM) and hyperglycemia

The metabolic effect of GDM (at diagnosis) and hyperglycemia (without the occurrence of GDM) have been studied through maternal urine analyzed by NMR (Sachse *et al.*, 2012) and maternal blood serum analyzed by MS (Scholtens *et al.*, 2013), respectively. Increased urinary citrate excretion, was found in urine of a large cohort of GDM-affected women at diagnosis and *pre*-diagnosis (Sachse *et al.*, 2012). The authors suggested that citrate increased was related with maternal hyperglycemia but concluded that the multivariate model could not reliably identify GDM cases (Sachse *et al.*, 2012). Recently, maternal hyperglycemia (without GDM) was associated with higher maternal plasma triglycerides, ketone bodies and amino acids, and low 1,5-anhydroglucitol, a naturally occurring dietary polyol (Dungan *et al.*, 2006), and a known marker of short-term glycemic control and predictor of fetal macrosomia in the 3rd trimester (Scholtens *et al.*, 2013). The quest for potentially predictive metabolite signature of GDM in 2nd trimester AF (Graça *et al.*, 2010, Graça *et al.*, 2012, Graça, 2013) and maternal urine (Diaz and Pinto *et al.*, 2011, Graça *et al.*, 2012, Diaz *et al.*, 2013) (two of these publications resulting from this work) has also been reported. The overall findings suggest early signs of maternal hyperglycemia, changes in amino acids metabolism, altered biotin status, nucleotide metabolism and gut microflora preceding the clinical manifestation of the disease (Diaz and Pinto *et al.*, 2011, Graça *et al.*, 2012, Diaz *et al.*, 2013). Regarding AF a subtler

metabolic effect was registered, the main difference being glucose increase, possibly caused by maternal hyperglycemia (Graça *et al.*, 2010, Graça *et al.*, 2012, Graça, 2013).

The impact of GDM in the newborns has also been studied through umbilical cord serum of newborns born to GDM mothers (Dani *et al.*, 2013). The authors found evidence of fetal hyperinsulinemia and increased oxidative stress (a known consequence of GDM), enhanced fetal glycolysis and perturbation in amino acids metabolism (consistently with findings in maternal biofluids) (Dani *et al.*, 2013).

Chromosomal disorders (CD)

A few studies have addressed the general effect of CD, by considering heterogeneous groups but also specific changes of trisomy 21 (T21), or Down's Syndrome (DS), and for trisomy 18 (T18). Reports on AF of an heterogeneous group of CD have unveiled significant changes in sugar metabolism (through the involvement of glucose and *myo*-inositol) and lipids and membranes synthesis (suggested by changes in *myo*-inositol). Maternal urine (Diaz and Pinto *et al.*, 2011, Diaz *et al.*, 2013) (publications comprising the results presented in this thesis) and blood serum have also been explored as a possible tool for the identification of T21 and T18 as well as to differentiate between these two CD types (Bahado-Singh *et al.*, 2013, Bahado-Singh *et al.*, 2013). T21 has been related to significant differences in 3-hydroxyisovalerate (3-HIVA), 3-hydroxybutyrate (3-HBA), suggestive of perturbations in brain growth and myelination and oxidative stress. The authors reported that the model considering these metabolites yielded a AUC of 0.757 whilst a combination of the same metabolites and maternal age had an AUC of 0.862 (Bahado-Singh *et al.*, 2013). Shortly after the same research group published a subsequent report for the detection of T18 ($n=18$ cases), stating significant changes in 2-hydroxybutyrate and glycerol, and that the combination of these metabolites and maternal age had an AUC of 0.92. Furthermore, it was found that T21 and T18 samples were distinguishable based on trimethylamine (TMA) contents (Bahado-Singh *et al.*, 2013).

Disorders of fetal growth (IUGR, SGA, LBW and VLBW)

The metabolic effect of fetal growth perturbations has also been addressed through metabolomic studies of maternal, fetal and newborn biofluids (Horgan *et al.*, 2010, Horgan *et al.*, 2011, Favretto *et al.*, 2012, Ivorra *et al.*, 2012, Alexandre-Gouabau *et al.*, 2013, Diaz

et al., 2013), one of which resulted from the work here presented. Altered amino acid transfer into fetal circulation and fetoplacental perfusion along changes in placental lipids metabolism has been reported in SGA cases (Horgan *et al.*, 2010, Horgan *et al.*, 2011). This altered placental function affects the metabolic symbiosis of mother and fetus which has been found in recent reports through the comparison of cord plasma of LBW (Ivorra *et al.*, 2012), VLBW cases (Alexandre-Gouabau *et al.*, 2013) and umbilical vein serum in cases of IUGR (Favretto *et al.*, 2012). Impaired fetal growth appears to be related with alterations in lipids, fatty acid oxidation, proteins metabolism, DNA methylation, polyamine flux and puridine metabolism in the fetus. Moreover, newborns' urine of IUGR babies confirmed perturbations in lipids and proteins metabolism along with evidence of altered insulin secretion (Dessi *et al.*, 2011).

Asphyxia, respiratory distress syndrome (RDS) and meconium aspiration syndrome (MAS)

The effect of neonatal respiratory complications, in newborn urine composition has been explored, with changes in urinary lactate, energy metabolism and organic acids apparently characterizing these conditions (Ma *et al.*, 1995, Chu *et al.*, 2006). In addition, umbilical cord blood of hypoxic ischemic encephalopathy (HIE, a complication of neonatal asphyxia) and asphyxia without HIE cases have been recently reported. The disorder was related with a derangement in energy metabolism arising from neonatal asphyxiation and neurological insult, up-regulation of triglyceride, phospholipids catabolism and ketone bodies production (Walsh *et al.*, 2012, Reinke *et al.*, 2013). Moreover, the authors reported that HIE could be detected based on 5 MS detected metabolites (decenoyl-L-carnitine, 3,5-tetradecadienecarnitine, phosphocholine C38:0, phenylalanine and proline) with AUC of 0.92 (Walsh *et al.*, 2012). By using an NMR model, 4 metabolites (3-hydroxybutyrate glycerol, *O*-phosphocholine and succinate) were found to discriminate HIE with an AUC of 0.80 and different grades of HIE severity with AUC of 0.88 to 0.95 (Reinke *et al.*, 2013).

Premature rupture of the membranes (PROM)

The unveiling of potentially predictive markers of PROM has also been attempted by analyzing AF (Graça *et al.*, 2010, Graça *et al.*, 2013), maternal blood plasma and urine

(Diaz and Pinto *et al.*, 2011, Diaz *et al.*, 2013) (the latter publications comprising results presented in this thesis) collected at the 2nd trimester (i.e. in the *pre*-diagnosis state). Only faint changes were identified in AF, these comprising proteins, amino acids, sugars and lactate, and in maternal blood, namely acetate, glutamine, citrate and albumin, whilst no change was found in maternal urine (Graça *et al.*, 2010, Diaz and Pinto *et al.*, 2011, Diaz *et al.*, 2013, Graça *et al.*, 2013). The low impact of PROM in all maternal biofluids studied can be a) a direct reflection of the low gravity of PROM (after the 37th g.w.), as it is usually resolved by inducing labor without posing any risk for the mother and fetus, and/or b) the high interval from sample collection to diagnosis.

***Other relevant studies (poor pregnancy outcome and effect of delivery mode),
Inborn Errors of Metabolism (IEM) and infections***

Recently, the possibility to predict poor pregnancy outcomes was attempted by comparing maternal blood serum samples collected during the 3rd trimester (Heazell *et al.*, 2012). However, in this study the poor pregnancy outcome group consisted in cases of SGA, *pre*-PTD and/or neonatal admission samples concomitantly, thus providing a general effect rather than disease-specific.

The metabolic effect of delivery mode, i.e. cesarean section (CS) *vs.* vaginal delivery (VD) has also been studied through the comparison of umbilical cord blood and placental extracts (Dunn *et al.*, 2012, Hashimoto *et al.*, 2013). The authors reported changes consistent with labor-imposed oxidative stress, which is up-regulated by hormonal action and enhanced in VD (Dunn *et al.*, 2012, Hashimoto *et al.*, 2013).

In relation to inborn errors of metabolism (IEM), newborn' dried blood spots, blood and urine have been studied. The first study searched for markers of phenylketonuria (PKU, a condition known to cause increased levels of phenylalanine in urine and blood) in newborns blood spots and urine. PKU was related to higher levels of phenylalanine in both biofluids (as expected), but no further interpretation on the disease was advanced (Constantinou *et al.*, 2004, Constantinou *et al.*, 2005). A recent study compared the levels of 400 compounds in newborns' dried blood spots of 21 cases of PKU, 21 cases of medium chain acyl-coenzyme A dehydrogenase deficiency (MCADD), and 24 other IEM. The study was carried out using a novel nanospray high resolution MS (nS-HR-MS) method,

the authors concluding that the proposed methodology is suitable for neonatal screening of IEM (Dénes *et al.*, 2012).

Metabolomics of newborn urine has been applied for the study of newborns' infection, namely for bacterial and CMV infections. Whilst no changes were found in bacterial infection, newborns infected with CMV were associated with altered energy metabolism along with other changes that the authors attributed to be virus related (Trump *et al.*, 2006, Fanos *et al.*, 2013).

Healthy pregnancy studies

Metabolic profiling studies of healthy pregnancies have been published in relation to gestational age and maternal nutritional states (Ottolenghi *et al.*, 2010, Athanasiadis *et al.*, 2011, Diaz *et al.*, 2012, Dunn *et al.*, 2012, Su *et al.*, 2012), one comprising results presented in this thesis. The first publications in relation to gestational age found that succinic acid and amino acids contents in AF varied as a function of gestational age (Ottolenghi *et al.*, 2010, Athanasiadis *et al.*, 2011). Another study compared placental extracts collected at early and late 1st trimester, and found significant changes in lipids composition, suggesting changes in fatty acids β -oxidation for mitochondrial ATP production within the 1st trimester (Dunn *et al.*, 2012).

Maternal serum metabolomic of healthy pregnant women living in a region with high prevalence of congenital anomalies has also been studied (Su *et al.*, 2012). The authors found a different nutritional and metabolic status in women living in the high prevalence area, highlighting a possible role of a panel of nutrients in the prevention/development of certain pregnancy complication (Su *et al.*, 2012).). A similar approach was used as a mean to identify the effect of pesticide exposure in pregnant women measured through maternal urine (Bonvallot *et al.*, 2013). This recent study compared urine from healthy pregnant women living in an agricultural region to evaluate the health risk associated with pesticides exposure. The authors concluded that higher oxidative stress and disturbed energy metabolism, occurs in higher exposures, a condition that could possibly result in disruptions to placental exchanges (Bonvallot *et al.*, 2013).

1.7 Aims and scope of this thesis

This work has entailed five general goals:

- The first was to establish the dynamic metabolic profile of maternal urine characteristic of healthy pregnancies, to unveil the control urinary trajectory describing each pregnancy trimester. This information would set the basis for the understanding of metabolic deviations in maternal urine composition caused by pregnancy disorders.
- The second goal was to evaluate the value of urine metabolomics to unveil potential biomarkers (either individual metabolites or metabolite profiles) for the recognition of FM and CD disorders. This study aims at improving the diagnostic methods presently available.
- The third goal was to search for potentially predictive signatures in 2nd trimester maternal urine for disorders developed later in pregnancy, particularly for PTD, GDM, PE, IUGR and PROM disorders (for which there are no predictive methods).
- The fourth goal was to deepen the understanding of the maternal metabolism disturbances caused by GDM and to understand the metabolic effect of treatment. Also, this work aimed at unveiling a potential *pre*-treatment metabolic phenotype predictive of future treatment requirements, towards the development of individualized healthcare strategies.
- The fifth and final goal was to unveil the metabolic impact of prenatal and newborn disorders in the babies' urine profiles in the first days of life.

2. Experimental section

This research was carried out with the collaboration of the Maternity Bissaya Barreto (MBB), Centro Hospitalar de Coimbra, where all samples were collected, from 2006 to 2012, and the Faculty of Medicine of the University of Coimbra (FMUC) where samples were stored immediately after collection. Between January 2008 and June 2011, this work was supported by the Portuguese Foundation of Science and Technology (FCT) funding through the PTDC/QUI/66523/2006 project. Samples were collected under ethical committee approval (Ref. 18/04 and 29/096) and informed consents were obtained, from all subjects participating in the study (see Annex II for informed consents and individual questionnaires).

2.1 Sampling

2.1.1 Sample and metadata collection

A summary scheme of the experimental design is shown in Figure 2.1, with indication of the entities involved in this research project. For pregnant women, samples were collected during medical appointments, during the morning period, under non-fasting conditions due to dietary restrictions in controlling/limiting pregnant women's diet. Approximately 50 mL of pregnant women urine were collected and stored at -20° for up to 2 hours at MBB. Newborn urine was collected during their stay at the MBB (day 1 to day 6 of life), after parental consent (Annex II). Samples were collected non-invasively using the cotton-ball method. This method consists in placing a sterile cotton-ball inside the newborns diaper, for up to 3 hours, to retain urine (approximately 1.5 mL) which is further transfer to an sterile vial and frozen at -20°C for up to 24 hours. All samples were then transferred from MBB to FMUC, where they were given a random code (to ensure patients confidentiality), frozen and stored at -80°C. Samples were then transported, in dry ice, to CICECO-Chemistry Department, at the University of Aveiro (UA) where they were stored and analyzed.



Figure 2.1: Experimental layout of the research project showing the involved research entities and list of urine sample collection, logistic and storage conditions for pregnant, non-pregnant and newborns. CICECO: Centro de Investigação em Materiais Cerâmicos e Compósitos. CHC: Centro Hospitalar de Coimbra, FMUC: Faculty of Medicine, University of Coimbra, UA: University of Aveiro.

Clinical and obstetrical metadata regarding prenatal and pre-pregnancy health history was obtained through individual questionnaires filled at the time of collection and at the end of gestation for all pregnancies. Other relevant parameters were obtained, such as height, pre-gravid weight and pre-gravid body mass index (BMI), preexisting disorders, medications, lifestyle and nutritional habits, smoking habits, blood pressure history, blood chemistry and urine analysis, obstetric background, fetal karyotype and ultrasound monitoring results. Information about labor (or pregnancy termination for cases of FM or CD with poor prognostic) and neonatal outcome was also obtained for each sample. Based on the collected metadata, samples were classified into controls, if they were obtained from healthy pregnant women with a normal pregnancy course and no preexisting maternal pathologies. Criteria for the definitions of other groups are described throughout this section.

2.1.2 Definition of sample groups

This thesis is divided into four main studies described in chapters 3, 4, 5 and 6. *Study 1* (Chapter 3) aimed the definition of the dynamic urinary profile of healthy pregnancies and the unveiling of the control urinary metabolic trajectory descriptive of each pregnancy trimester. In *Study 2* (Chapter 4), the possibility of using 2nd trimester maternal urine for the diagnosis of fetal disorders and prediction of poor pregnancy outcomes was explored. *Study 3* (Chapter 5) aimed at the understanding of the metabolic alterations in GDM by characterizing the urinary profile at three disease stages (compared to controls) as well as improving the knowledge of the metabolic effect of GDM treatments and an attempt to predict future treatment requirements. *Study 4* aimed at the unraveling of the potential of newborn urine metabolomics for the understanding of the metabolic impact of prenatal disorders in newborns' health. A schematic representation of the a) disorders investigated in each study and b) gestational times at sampling is shown in Figure 2.2.

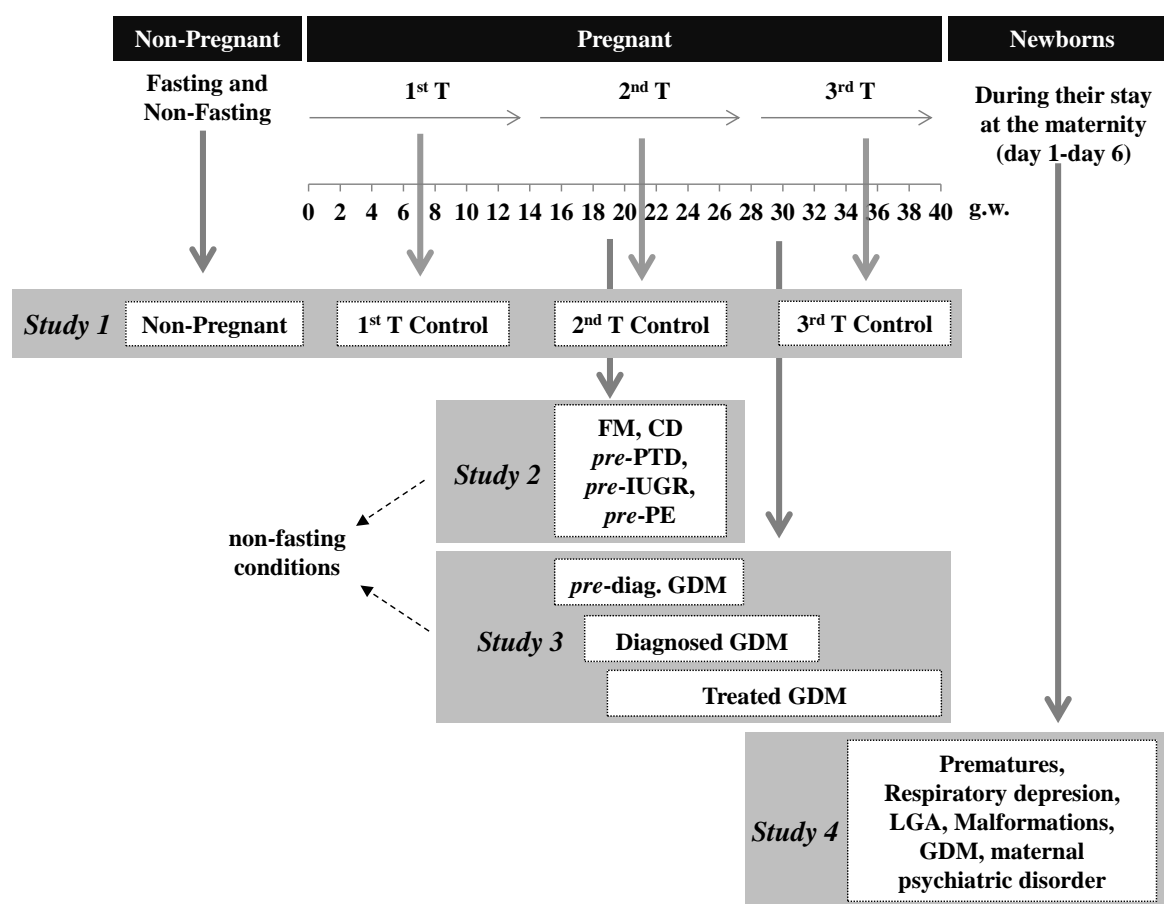


Figure 2.2: Schematic representation of the disorders investigated in *Studies 1 to 4* and corresponding gestational times at sampling.

Sample groups in Study 1: Defining the global urinary profile of pregnancy and the metabolic adaptations throughout gestation - a multipoint collection study

For the study of the urinary profile of healthy pregnancies and the metabolic trajectory throughout gestation only control cases were used. Urine samples were collected for independent groups of subjects at their 1st, 2nd and 3rd trimesters, the number of samples for each independent subject group, along with corresponding metadata information being listed in Table 2.1.

For this study, independent groups were chosen, rather than multiple-collection for each subject, in order to consider pregnancy-related changes larger than inter-subject variability, reflecting the general population more closely. For a group of non-pregnant (NP) women, urine was collected fasting and non-fasting (2 hours after random breakfast), in order to evaluate the effect of the typical non-fasting conditions in which sample collection took place for pregnant women.

Table 2.1: List of urine samples collected for *Study 1*, along with corresponding ranges of maternal age (in years), pre-pregnancy body mass index (BMI, in kg.m⁻²), gestational age at sampling and at delivery (in gestational weeks, g.w.), gravidity (no. pregnancies), parity (no. born children), newborn's birth weight (kilograms, kg), no. pregnant women who were smokers and who were taking folic acid supplementation at the time of sampling. NP: non-pregnant. Median values are shown in brackets. NA: Not available.

Metadata	Groups studied			
	NP	1 st T	2 nd T	3 rd T
<i>n</i>	16	16	20	19
Maternal age range	21-47 (26)	19-40 (31)	21-39 (34)	20-38 (32)
<i>pre</i> -gravid BMI range	18-35 (22)	19-25 (24)	20-33 (24)	17-35 (23)
Gestational age at sampling/g.w.	-	11-13 (12)	15-26 (17)	29-39 (35)
Gestational age at delivery/g.w.	-	37-40 (40)	37-40 (38)	37-40 (38)
Gravidity/Parity	-	1-2(1)/0-1(0)	1-6(2)/0-2(1)	1-5(2)/0-3(0)
Newborn's birth weight /kg	-	2.59-4.11 (3.40)	2.61-3.74 (3.24)	2.33-3.95 (3.27)
No. smokers	NA	0	1	2
No. subjects taking folic acid at sampling	NA	12	13	17

Sample groups used in Study 2: Second trimester maternal urine for the diagnosis of fetal disorders and prediction of poor pregnancy outcomes

Maternal urine was collected for women carrying fetuses diagnosed with fetal malformations (FM) and chromosomal disorders (CD) and for other women with no apparent disorder. Subjects were followed until the end of gestation and pathological groups were defined as: women that further developed preterm delivery (*pre*-PTD), preeclampsia (*pre*-PE), intrauterine growth restriction (*pre*-IUGR) and premature rupture of the membranes (*pre*-PROM), as no sufficient number of samples were gathered for other relevant disorders. Sample numbers, maternal age, gestational ages at sampling and *pre*-pregnancy BMI are listed in Table 2.2.

Table 2.2: Maternal urine samples collected for *Study 2*, along with corresponding age/years and gestational age/gestational weeks, g.w. at the time of collection, *pre*-pregnancy body mass index, BMI/kg.m⁻² and, for *pre*-diagnosed disorders, the no. of g.w. between sampling and time of diagnosis. Values in brackets correspond to median values.

Subjects group	<i>n</i>	Maternal age /years	Gestational age /g.w.	Pre-pregnancy BMI / kg.m ⁻²	g.w. to diagnosis
Controls	84	21-42 (36)	15-26 (17)	17-35 (22)	—
FM	35	18-45 (31)	14-24 (21)	18-35 (25)	—
CD	33	25-47 (36)	15-25 (17)	20-35 (24)	—
<i>pre</i> -PTD	26	26-41 (36)	16-21 (17)	19-40 (25)	11-20 (17)
<i>pre</i> -PE	9	24-41 (35)	16-17 (17)	19-33 (24)	14-21 (18)
<i>pre</i> -IUGR	10	29-42 (39)	16-24 (17)	19-33 (24)	2-22 (19)
<i>pre</i> -PROM	68	22-44 (36)	15-21 (17)	19-36 (23)	18-24 (22)

FM cases were diagnosed by morphological ultrasound after 12 g.w., the group comprising cases of central nervous system (CNS), abdominal, urogenital, cardiac, limbs and poli-malformations and CD diagnosed *ca.* 2 weeks after amniocentesis (thus after sample collection), comprising trisomies 18 and 21 (T18 and T21), Klinefelter syndrome, mosaicisms, chromosome inversions and translocations. The numbers and specific types of FM and CD types are summarized in Table 2.3, showing the heterogeneity nature of these groups.

For undiagnosed later-developing disorders group cases were defined as follows: *pre*-PTD, women that later delivered before the 37th g.w.; *pre*-PE, women later diagnosed with PE according to the criteria of blood pressure $\geq 140/90$ mmHg after the 20th weeks of

gestation and proteinuria $\geq 0.3\text{g/L}$ in 24-hour urine, *pre*-IUGR, corresponding to women later diagnosed with carrying a IUGR fetus according to the diagnostic criteria of fetal growth below the 5% percentile, and *pre*-PROM, women that had premature rupture of membranes prior to labour and after the 37th g.w.

Table 2.3: List of specific types of FM and CD along with corresponding number of samples.

Fetal Malformations (FM)			
Type	<i>n</i>	Type	<i>n</i>
CNS		Limbs	
Corpus callosum agenesis	3	Club feet	3
Ventriculomegaly	4	Jaw malformation + club foot	1
Ventriculomegaly + polydactyly	1	Limb agenesis	1
Spina bifida	1	Polydactyly	1
Hydrocephaly	1	Cardiac	
Arnold-Chiari	1	Complex cardiopathy	5
Holoprosencephaly	1	Hypoplastic left heart syndrome	1
Urogenital		Polimaformation	
Bladder agenesis	1	Cystic hygroma + bladder agenesis	1
Hydronephrosis	2	Cystic hygroma + sexual ambiguity	1
Multicystic dysplastic kidney	1	Dextrocardia + lung agenesis + lung hypoplasia	1
Polycystic kidney	1	Hypertelorism+ club feet + lung ectasia + cardiac septum edema + liver fibrosis	1
Abdominal			
Enlarged stomach	2		
Chromosomal Disorders (CD)			
Type	<i>n</i>	Type	<i>n</i>
Trisomy 21	13	Mosaicism	2
Trisomy 18	2	Marker chromosome	1
Klinefelter syndrome	2	Translocations	5
XYY syndrome	2	Inversions	2
XXX syndrome	1	Deletions	2
Triploidy	1		

Sample groups in Study 3: Gestational Diabetes Mellitus - a deeper insight into maternal metabolism disturbances

For the study of the metabolic perturbances in GDM-affected pregnancies, urine samples were collected from independent groups of pregnant women at 2nd and 3rd trimesters, during routine medical appointments and diabetes appointments. GDM was diagnosed after a 75g or 100g oral glucose tolerance test (OGTT) by measuring plasma

glucose (PG) at fasting (FPG) and after 1, 2 and 3 (for the latter) hours. GDM was diagnosed according to the diagnosis criteria described in Table 1.7.

Disorder groups correspond to samples collected at three different disease stages:

- *Diagnosed GDM:* samples collected at the first medical appointment after GDM diagnosis (2nd and 3rd trimester) and before the initiation of any treatment.
- *Pre-diagnosis GDM:* samples collected in the 2nd trimester, before any clinical manifestation of the disease.
- *Treated GDM:* samples collected at follow up medical appointment (2nd and 3rd trimester), upon treatment course (1600-2000 kcal diet further complemented with insulin therapy). Insulin was prescribed to pregnant women with FPG \geq 5 mmol/L and/or post-prandial, one hour after main meals, PG \geq 6.65 mmol/L after initiation of dietary intervention.
- *Control groups 1, 2 and 3:* Due to the significant impact of gestational age on the composition of urine, during pregnancy (as will be discussed after in chapter 3), three control groups were defined, based on gestational age range, in order to match controls to the corresponding disorder group (*diagnosed GDM*, *pre-diagnosis GDM* and *treated GDM*, respectively), as far as possible.

Detailed information about samples number, maternal age range, gestational age range (at sampling), interval of time from sampling to diagnosis and pre-pregnancy BMI range can be found in Table 2.4.

Table 2.4: List of urine samples collected for *Study 3*, comprising each independent group of GDM-affected subjects (in three different disease stages) and controls (control groups 1, 2 and 3, to match each disorder group, as described in the previous paragraph), corresponding maternal age/years and GA/g.w.at sampling, interval between sampling and diagnosis/g.w. and *pre*-pregnancy BMI (kg.m⁻²). Median values are shown between brackets. GA: gestational age.

Group	<i>n</i>	Maternal age range/years	GA range/g.w.	g.w. to diagnosis	<i>Pre</i> -pregnancy BMI range/kg.m ⁻²
<i>Diagnosed GDM</i>	13	26-41 (34)	16-33 (27)	0-7 (2)	18-36 (25)
Control group 1	14	24-39 (32)	16-33 (31)	—	17-34 (23)
<i>pre-diagnosis GDM</i>	52	30-44 (37)	15-22 (17)	2-21 (13)	18-40 (25)
Control group 2	113	21-42 (36)	15-26 (17)	—	16-39 (22)
<i>Treated GDM</i>	38	18-41 (32)	17-40 (35)	2-22 (4)	18-45 (25)
Control group 3	40	20-41 (34)	16-39 (31)	—	17-34 (23)

Sample groups in Study 4: Newborns health status seen through urine metabolomics

Newborns groups were defined based on maternal and newborns' medical records. The control group was defined as healthy newborns, born to healthy mothers and pregnancies. For the control group, the effect of different variables which may play as confound in the later studies devoted to disorders, namely gender, delivery mode (vaginal vs. abdominal), newborns age in days (day 1 to day 4), and GA at birth (37 to 38, 38 to 39, 39 to 40 and over 40 g.w.). Sample numbers for these possibly confounding effects are listed in Table 2.5 (bottom).

Table 2.5: List of newborn urine samples collected *Study 4*, along with corresponding day of life at sampling/days, gestational age range/g.w., birth weight range/kg, and maternal age range/years (top); and list of possible confounding effects and corresponding sample numbers (considered only for the for the control group (bottom)). Median values are shown between brackets. LGA: large for gestational age, Mat: Maternal, F: female, M: male, VD: vaginal delivery, CS: cesarean section.

Group	<i>n</i>			Day of life at collection	GA/g.w.	Birthweight /kg	Mat. age /years	Gravidity/ Parity
	Total	M/F	VD/CS					
Control	45	23/23	29/17	1-4 (2)	37-40 (39)	2.46-3.69 (3.11)	16-43 (31)	1-6 (2)/ 0-3 (1)
PROM	36	22/14	21/15	1-3 (2)	35-40 (39)	2.23-4.16 (3.23)	19-41 (33)	1-4 (1)/ 0-3 (1)
LGA	18	14/4	7/11	1-2 (2)	38-40 (39)	3.70-4.76 (3.92)	20-38 (34)	1-4 (2)/ 0-3 (1)
Prematurity	17	11/6	8/9	1-6 (2)	33-36 (35)	1.26-3.28 (2.36)	23-41 (30)	1-3 (1)/ 0-2 (0)
<i>pre</i> -jaundice	12	8/4	8/4	1-3 (2)	35-40 (38)	2.16-3.47 (2.99)	24-41 (33)	1-3 (1)/ 0-1 (0)
Respiratory depression	10	4/6	6/4	1-2 (2)	36-40 (39)	2.49-3.49 (3.13)	20-36 (33)	1-4 (2)/ 0-1 (1)
Malformations	9	6/3	2/7	2-3 (2)	37-40 (39)	2.61-3.59 (3.02)	19-40 (33)	1-4 (2)/ 0-3 (1)
Newborns to GDM mothers	14	8/6	6/8	1-6 (2)	33-40 (39)	1.27-4.77 (3.28)	25-40 (32)	1-6 (4)/ 0-3 (1)
Mat.psychiatric disorders	7	2/5	3/4	1-2 (2)	38-40 (39)	2.44-4.06 (3.26)	20-39 (35)	1-5 (2)/ 0-2 (1)
Mat. respiratory disease	10	5/5	7/3	1-3 (2)	35-40 (38)	2.64-4.16 (2.88)	23-36 (32)	1-4 (2)/ 0-3 (0)
Mat. thyroid disease	9	6/3	6/3	1-2 (1)	34-40 (39)	1.98-3.22 (2.88)	20-37 (32)	1-4 (2)/ 0-3 (1)
Mat. chronic hypertension	8	3/5	5/3	1-6 (2)	33-39 (37)	1.27-4.10 (2.61)	25-44 (32)	1-4 (2)/ 0-3 (1)
<i>Confounding effects investigated</i>								
Gender	F=23		M=23					
Day of life	Day 1= 7		Day 2=31		Day 3=2	Day 4=2	NA=4	
GA/g.w.	37≤ GA <38=5		38≤ GA <39=10		39≤ GA <40=18	40≤ GA <41=13		
Delivery mode	VD= 29		CS=17					

Pathological cases were defined for a) newborns disorders, these including cases large for gestational age (LGA, birth weight above the 90th percentil), prematurity (babies born before the 37th g.w.), *pre*-jaundice (babies later developing jaundice), respiratory depression (babies requiring reanimation with external O₂) and malformations (heterogeneous group, 1 major malformation and 8 minor malformations) and b) newborns from pathological pregnancies, including newborns born to GDM mothers, maternal psychiatric disorders (including depression, epilepsy and depressive neurosis), maternal respiratory disease (including asthma, bronquitis and sinusitis), maternal thyroid disease (including hypo-thyroidism and other thyroid diseases), maternal hypertension and babies born from pregnancies affected by PROM. Sample numbers, maternal age, day of life at sampling, gestational age at birth (g.w.), birth weight, maternal age, gravidity and parity are listed in Table 2.5 (top).

2.2 NMR spectroscopy

2.2.1 Sample preparation for NMR analysis

Samples were thawed at room temperature and 800 μ L (for pregnant women) and 600 μ L (for newborn urine) were centrifuged (4500g, 5 min). Then, 60 μ L of 1.5M KH₂PO₄/D₂O phosphate buffer pH 7.00, 0.1%Na⁺/3-trimethylsilylpropionic acid (TSP) were added to 540 μ L of supernatant, followed by buffering and pH adjustment to 7.00 \pm 0.02 with KOD (4M) or DCl (4M). The mixture was again centrifuged (4500g, 5 min) and 550 μ L were transferred to a 5 mm NMR tube. The initial, that after buffering and final pHs are listed in Table Table 2.6.

2.2.2 Acquisition and processing of NMR data

All 1D, TOCSY and HSQC spectra were recorded on a Bruker Avance DRX 500 spectrometer, operating at a proton frequency of 500MHz, and *J*-res were recorded in a 600MHz spectrometer Bruker BioSpin, Germany. All spectra were acquired at 300 K. Standard 1D spectra were acquired, using a *noesypr1d* 1D pulse sequence (RD-90°-t₁-90°-t_m-90°-acquire) experiment from Bruker library. This experiment performs water presaturation, using a presaturation pulse of 60 dB, during the relaxation delay (RD) and mixing time (t_m) allowing an efficient water suppression. Each FID was multiplied by 0.3

Hz exponential line-broadening followed by FT. Spectra were then manually phased and baseline corrected, and chemical shifts calibrated internally by TSP at $\delta=0.0$ ppm. A list of all acquisition and processing parameters is show in Table 2.7.

Table 2.6: Initial pH, pH after buffering and final pH (after addition of KOD or DCl) ranges, median values in brackets, of groups studied in this work.

Sample group	Intial pH	pH after buffering	Final pH
NP	5.06-7.90 (5.35)	6.51-6.96 (6.66)	6.98-7.02 (7.01)
1 st T controls	5.34-7.91 (6.35)	6.54-6.95 (6.84)	6.98-7.02 (7.00)
2 nd T controls	5.03-8.45 (6.67)	6.42-7.09 (6.87)	6.98-7.02 (7.01)
3 rd T controls	5.22-7.76 (6.31)	6.46-6.99 (6.81)	6.99-7.02 (7.01)
Pregnant women			
FM	5.14-8.56 (7.04)	6.56-7.06 (6.87)	6.98-7.02 (7.00)
CD	5.51-8.01 (6.82)	6.60-7.05 (6.89)	6.98-7.02 (7.01)
<i>pre</i> -PTD	5.24-8.37 (6.55)	6.59-7.02 (6.85)	6.99-7.02 (7.01)
<i>pre</i> -PE	5.25-7.84 (6.35)	6.72-6.97 (6.89)	6.99-7.02 (7.00)
<i>pre</i> -IUGR	5.36-7.20 (6.35)	6.70-6.99 (6.88)	6.98-7.02 (7.00)
<i>pre</i> -diagnosis GDM	5.08-8.11 (6.61)	6.50-7.00 (6.82)	6.98-7.02 (7.01)
Diagnosed GDM	5.63-7.55 (6.66)	6.53-6.92 (6.79)	6.98-7.02 (7.00)
Treated GDM	5.36-7.85 (6.52)	6.35-7.02 (6.87)	6.99-7.02 (7.01)
Newborn			
Controls	5.70-6.61 (5.99)	6.67-6.97 (6.85)	6.98-7.02 (7.00)
Prematures	5.75-7.15 (6.14)	6.80-6.98 (6.90)	6.98-7.02 (7.01)
Respiratory depression	5.64-7.43 (5.96)	6.77-6.99 (6.86)	6.99-7.02 (7.00)
LGA	5.55-6.47 (5.87)	6.60-6.91 (6.74)	6.99-7.02 (7.01)
Malformations	5.50-7.32 (5.90)	6.66-6.96 (6.85)	6.99-7.01 (7.00)
Maternal GDM	5.32-6.86 (5.97)	6.62-6.92 (6.85)	6.99-7.02 (7.00)
Maternal psychiatric disorders	5.55-6.81 (5.74)	6.60-6.98 (6.78)	6.98-7.02 (7.00)

The TOCSY experiment was recorded in phase-sensitive mode using States-TPPI (time proportional phase incrementation) detection in t_1 using a MLEV-17 pulse sequence (Bax and Davis, 1985), *clmlevprtp*. The HSQC spectra were recorded with inverse detection and ^{13}C decoupling during acquisition using the *invietgpsi* pulse program. *J-resolved* experiments were acquired using the *jresgppraqf* pulse sequence with presaturation during relaxation delay. All NMR data processing was carried out in TOPSPIN 2.0 and 3.2 versions (Bruker BioSpin, Reinstetten, Germany).

Table 2.7: List of acquisition and processing parameters used for 1D and 2D NMR experiments of urine. NS: number of scans, RD: relaxation delay, t_m : mixing time, t_1 : fixed delay, SW: spectral window, TD: size of FID, RG: receiver gain, SI: size of real spectrum, WDW: type of window function, F1: 1st dimension, F2: 2nd dimension, lb: line broadening, efp: exponential multiplication, Fourier transform and phase correction.

<i>Standard 1 D experiment (noesypr1d)</i>					
Acquisition parameters				Processing parameters	
NS	128			EFP and lb=0.3 Hz	
RD	4 s			Manual phase correction	
t _m	100 ms			Manual baseline correction	
t ₁	3μs			Chemical shift calibration by TSP (0.0 ppm)	
SW	10080.65 Hz/ 20.16 ppm				
TD	64 k				
RG	32				
<i>TOCSY experiment (clmlevprtp)</i>					
Acquisition parameters				Processing parameters	
NS	24	RD	2 s	SI [F1]	1 k
t _m	80 ms	pl10	15 dB	SI [F2]	4 k
TD [F1]	2 k	TD [F2]	300	WDW	qsine
SW [F1]	8012.82 Hz	SW [F2]	8012.82 Hz		
<i>HSQC experiment (invietgpsi)</i>					
Acquisition parameters				Processing parameters	
NS	36	RD	2 s	SI [F1]	1 k
TD [F1]	4 k	TD [F2]	300	SI [F2]	4 k
SW [F1]	25153.81 Hz	SW [F2]	8012.82 Hz	WDW	qsine
<i>J-resolved experiment (jresgpprqf)</i>					
Acquisition parameters				Processing parameters	
NS	1	RD	1 s	SI [F1]	128
TD [F1]	8192 Hz	TD [F2]	40	SI [F2]	8 k
SW [F1]	78.16 Hz	SW [F2]	10000 Hz	WDW	qsine

2.2.3 Preprocessing and multivariate analysis of NMR data

Alignment and normalization

NMR spectra were converted to a matrix of n rows (corresponding to n samples), and m columns (corresponding to variables). The NMR spectra, after removal of water (4.60-5.05 ppm) and urea (5.50-6.45 ppm) regions, were aligned using a recursive segment-wise peak alignment (Veselkov *et al.*, 2009) to minimize chemical shift variations and normalized by PQN (Dieterle *et al.*, 2006) using MATLAB software, version 7.12.0, The MathWorks Inc. The script was developed at Imperial College of London and kindly spared by Dr. Kirill A. Veselkov.

Variable selection for NMR data

Four variable selection methods were tested, based on the selection of variables (spectral data points) with $VIP > 1$ (Quintás *et al.*, 2012, Sun *et al.*, 2012), variables with lower standard errors in relation to VIP (VIP_{cvSE}) and to b-coefficients (b_{cvSE}) (Andersen and Bro, 2010), given respectively by $VIP/VIP_{cvSE} > 1$ and $b/b_{cvSE} > 1$ and, finally, variables selected through the intersection of all three previous methods, i.e. simultaneously with high VIP, low VIP_{cvSE} and low b_{cvSE} (Diaz *et al.*, 2013), this method being published in a paper originated from the present thesis. The standard error obtained by cross validation (cvSE) is based on the concept of signal to noise ratio using a jack-knife approach. Jack-knife during cross-validation is a mean to estimate statistical error by re-computing models using a random subset of samples (Andersen and Bro, 2010). Thus, the cvSE provides a measurement of the uncertainty of a parameter estimate and is helpful for the selection variables relevant to the model. This concept was applied to the b-coefficient and VIP parameters, chosen due to their complementarities. The b-coefficient reflects the discriminant and predictive capability of each variable i.e. their contribution for samples classification, whilst VIP values, which reflect the importance of each variable to the projection in the low dimensional space, hence, the relevance of each variable for samples geometric separation.

Multivariate analysis

MVA was carried out using SIMCA-P software, version 11.5, by applying PCA and PLS-DA to the UV scaled matrix. Model robustness was initially evaluated in terms of Q^2 , the goodness of prediction or prediction power, of the original model, that is of the model calculated in SIMCA using a 7-fold cross validation. The corresponding loading plots were back-transformed by multiplying the loading weight [**w**] by the variable standard deviation. Loading plots were colored according to the VIP value and relevant peaks were integrated in the original spectra using Amix 3.9.5, Bruker BioSpin, Rheinstetten, Germany.

Model Validation

Model validation was further assessed by MCCV (7 blocks) using 500 runs, with recovery of Q^2 values and confusion matrices of true and permuted classes. Classification

rates (CR), specificity and sensitivity were calculated from confusion matrices, and model predictive power was assessed using a ROC curve mapping, a function of the true positive rate (TPR or sensitivity) and false positive rate (FPR or 1-specificity). PLS-DA models were considered validated when having high predictive power and robustness, given by minimal overlapping of the distribution of true and permuted Q^2 , no overlapping of true and permuted models in the ROC plot and high CR, specificity and sensitivity (Wiklund *et al.*, 2007, Westerhuis *et al.*, 2008). MCCV calculations were carried out using the *spark* algorithm, developed and kindly shared by Dr. António S. Barros.

Univariate analysis

Integrals were normalized by PQN, i.e. divided by the each samples' normalization quotient in *Study 2 to 4* and by total intensity in *Study 1*. All integrals were compared by determining individual statistical significance, calculated by Wilcoxon rank sum test (considering a significance level of 95%, $p < 0.05$). For each set of metabolites, the biological relevance was determined by the effect size (considered in order to take into account sample dispersion and sample number) (Nakagawa and Cuthill, 2007). Effect size was calculated according to Eq. 2.1 and Eq. 2.2, where \bar{x}_1 and \bar{x}_2 are the average metabolite integrals for sample groups 1 and 2 respectively, s is the pooled variance between the two groups, n_1 and n_2 are the sample numbers respectively for groups 1 and 2, s_1^2 and s_2^2 are the and the variances respectively for groups 1 and 2.

$$effect\ size = \frac{\bar{x}_1 - \bar{x}_2}{s} \quad [Eq. 2.1]$$

$$s^2 = \frac{(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2}{n_1 + n_2 - 2} \quad [Eq. 2.2]$$

Metabolites variations were shown, for *Study 1*, using a heatmap representation, with integrals values normalized to unity and shown in a qualitative color scale. Statistical tests, heatmaps and VIP-coloured loadings plots were carried out using R-statistical software. In *Study 4*, each integrals' individual biological (effect size) and statistical (p -value) significance was further represented by means of a Volcano plot which is a graphical summary of the effect size and the p -value (Cui and Churchill, 2003).

2.3 Ultraperformance liquid chromatography- mass spectrometry (UPLC-MS)

UPLC-MS experiments were performed at the Section of Computational and Systems Medicine, Department of Surgery and Cancer of the Faculty of Medicine, Imperial College of London, within an ongoing collaboration between that institution and the University of Aveiro.

2.3.1 Sample preparation for UPLC-MS analysis

Samples were thawed at room temperature and centrifuged at 9500g for 10 min. Samples were then diluted (in a 1:1 ratio) by mixing 200 μ L of supernatant and 200 μ L of HPLC grade water and transferred to a 96-well plate (kept at 4°C until analysis). Quality control samples (QC) were prepared, by mixing equal volume of all diluted samples, to ensure the reproducibility throughout the run.

2.3.2 Acquisition and processing of UPLC-MS data

Urine samples were analyzed in an Acquity UPLC system (Waters Ltd. Elstree, UK) connected to an LCT Premier mass spectrometer (Waters MS Technologies, Ltd., Manchester, UK). A high strength silica (HSS, Acquity HSS T3 1.7 μ m, 2.1 x 100 mm, Waters Corporation, Milford, USA) column and a hydrophilic interaction chromatography (HILIC, Acquity BEH HILIC, 1.7 μ m, 2.1x100 mm, Waters Corporation, Milford, USA) column were used. Both columns were used due to their complementarity as HSS columns shows improved retention of more polar metabolites (than alternative traditional reverse phase columns (RP)) whilst HILIC columns provide improved means of profiling certain polar metabolites not easily retained by RP columns (Want *et al.*, 2010, Spagou *et al.*, 2011). Solvent gradients were used for both HSS and HILIC with a flow rate of 0.5mL·min⁻¹ and 0.4mL·min⁻¹ respectively. The corresponding gradients are shown in Table 2.8.

Table 2.8: Gradients used for HSS and HILIC columns. HSS solvent A- water with 0.1% formic acid, Solvent B- acetonitrile with 0.1% formic acid. HILIC solvent A-95% ammonium acetate 10 mM /5% acetonitrile with 0.1% formic acid, Solvent B- 50% ammonium acetate 10mM/50% acetonitrile with 0.1% formic acid.

Gradients used for HSS			Gradients used for HILIC		
	A _{HSS} (%)	B _{HSS} (%)		A _{HILIC} (%)	B _{HILIC} (%)
t=0	100	0	t=0	99	1
t=1	100	0	t=1	99	1
t=3	85	15	(linearly decreasing	0	100
t=6	50	50	until 12 min)		
t=9	5	95	t=12	99	1
t=10	5	95	t=15	99	1
t=10.1	100	0			
t=12	100	0			

Mass spectrometry (MS) analysis was performed in both positive and negative electrospray ionization modes (respectively ESI+ and ESI-), the instrument's parameters being summarized in Table 2.9. The MS was operated in V optics mode with scan time 0.1s and interscan delay of 0.01s, and data was collected in the range of 50-1000 m/z in centroid mode. Leucine enkephalin (a peptide with monoisotopic mass 555.2645 Da, commonly used in MS as a reference compound) was used as lock mass (solution of 200 pg μL^{-1}) by spraying the solution into the instrument at a rate of $3\mu\text{L}\cdot\text{min}^{-1}$, with lock mass scans collected every 15s and averaged over 3 scans to perform mass correction. 10 conditioning QC were run at the beginning of each experiment, and repeated every 10 samples to check for reproducibility within runs. Cases and control samples were run in a randomized order.

Table 2.9: MS spectrometer acquisition parameters for ESI+ and ESI- modes.

	Instrument parameters	
	ESI+	ESI-
Capillary voltage	3000 V	2500 V
Sample cone	30 V	25 V
Dessolvation temperature	400 °C	350 °C
Source temperature	120 °C	120 °C
Dessolvation gas flow	600 L h ⁻¹	800 L h ⁻¹

2.3.3 Preprocessing and multivariate analysis of UPLC-MS data

Raw data was converted to netCDF format using Waters MassLynxTM software version 4.1 Databridge. Chromatograms were converted into a data table using R-statistical software, version 2.15.0, along with the XCMS package (Smith *et al.*, 2006). Processing scripts were developed in Imperial College of London and kindly spared, courtesy of Dr. Mathew Lewis and Dr. Paul Benton. In this work, two preprocessing methods were employed as described below.

UPLC-MS data processing method 1:

Peak detection was applied using the peak picking CentWave algorithm with a peak window of 2-15 s for HSS and 3-60 s for HILIC columns. Retention times (RT) were corrected by applying a first peak grouping using an m/z bandwidth of 0.05 Da and RT error of 5 s, followed by a second grouping, this time using a RT error value determined from the RT deviation plot. Features were normalized using the median fold change normalization to account for differences in samples dilution (Veselkov *et al.*, 2011). An initial PLS-DA was computed and variables (in this case, features, i.e. a pair RT_m/z) were selected using the variable selection method used for NMR data (i.e. simultaneously $VIP > 1$, $VIP/VIP_{cvSE} > 1$ and $|b_{coeff}/b_{coeffcvSE}| > 1$), previously described in section 2.3.

UPLC-MS data processing method 2:

UPLC-MS data were subjected to an initial prefilter, to remove features that are not detected in at least 5 scans with intensity higher than 1000, followed by peak picking as described in *processing method 1*. The prefilter prior to peak peaking is extremely useful in removing noise and background from the dataset and consequently reducing their influence of the definition of features. Features were then grouped, also as described in *processing method 1*. A following MinFrac Filter was applied, this to remove features not present in at least 50% of at least one of the groups. Features were normalized as in *processing method 1*. Finally, a coefficient of variation (Eq. 2.3) based filter is also applied, with features with coefficient of variation higher than 30% (calculated for the QCs) were removed.

$$\text{Coefficient of variation} = \frac{\text{standard error}}{\text{mean}} \times 100 \quad [\text{Eq. 2.3}]$$

Multivariate analysis, model validation and univariate analysis of UPLC-MS data

The resulting data tables were analyzed by applying PCA and PLS-DA, using SIMCA-P version 11.5, combined with UV and Pareto scaling and logarithmic transformation. Reproducibility was confirmed for every dataset through PCA of cases and QC samples. Model validation was performed by MCCV, as for the NMR results. For validated models, features were analyzed by the S-plot, which combines covariance and correlation of each feature with the scores. Features with covariance or $|p| > 0.02$ and correlation or $|p_{\text{corr}}| > 0.6$ were selected for univariate analysis. As for NMR data, individual significance of feature intensities was calculated by Wilcoxon test and, for each set of significant features ($p < 0.05$), the effect size (Nakagawa and Cuthill, 2007) was determined, according to Eq. 2.1 and Eq. 2.2..

2.4 Statistical correlation analysis

Statistical total correlation spectroscopy (STOCSY) was performed for each unassigned NMR resonance in order to identify peaks arising from the same molecule and/or metabolically related compounds. Correlation and covariance was calculated between each unidentified peak and the complete spectrum matrix. The correlation coefficient (r) was then plotted in the covariance map, 1D STOCSY, (Cloarec *et al.*, 2005) without any threshold to facilitate the identification of correlated peaks.

For unidentified MS features, statistical heterospectroscopy (SHY) was applied by calculating the correlation between each MS feature and the full NMR matrix. The correlation coefficient was computed using Pearson correlation and plotted in a 1D STOCSY, i.e. using the covariance map (Cloarec *et al.*, 2005). A correlation and significance threshold of $|r| > 0.8$ and significance $p < 0.01$ were applied. All correlations were checked by scatter plotting every pair of metabolite integrals/features intensities to confirm validity and rule out any spurious correlation.

3. Defining the global urinary profile of pregnancy and the metabolic adaptations throughout gestation - a multipoint collection study

This chapter begins by presenting the results obtained for the characterization of maternal urine composition through 1D and 2D NMR experiments and the resulting complete list of metabolites identified. Secondly, the results obtained for the impact of non-fasting conditions as a possible confounding effect in urine metabolomic studies are presented. Finally, the overall changes found in maternal urine composition with and throughout healthy pregnancies are presented, with results of the metabolic profiling of urine for each pregnancy trimester and suggestions of possible metabolic interpretations.

3.1 Typical ^1H NMR spectrum of maternal urine and spectral assignments

Generally, a typical ^1H NMR spectrum of urine contains hundreds of peaks arising from numerous metabolites. As urine is mainly composed of small molecules originating sharp peaks, only 1D standard (*noesypr1d*) NMR spectra were acquired, as CPMG or diffusion-edited experiments were not justified. However, the peaks observed in the standard spectra overlap extensively, yielding complex spectra and hindering metabolite assignment. Typical ^1H NMR urine spectra of healthy pregnant women, in their 1st, 2nd and 3rd T, and that of NP women are shown in Figure 3.1. In this figure, the large complexity and extensive peak overlap are clearly visible, justifying the absolute need for 2D NMR. Thus, 2D NMR spectra TOCSY, HSQC and *J*-res were used and expansions of a TOCSY and HSQC spectra of urine (0.5-4.6 ppm) are shown in Figure 3.2 with indication of some assignments. TOCSY is extremely helpful for the assignment of complete spin systems, especially when peaks are present in low intensities and often obscured in the 1D spectrum due to overlap with other resonances. On the other hand, the HSQC is particularly helpful for the assignment of singlet resonances, which would have no scalar correlation and are therefore not observed in TOCSY spectra.

Other relevant assignment tools comprise statistical TOCSY (STOCSY) and the traditional spiking experiments, an example of each procedure being shown in Figure 3.3a,b respectively. In the STOCSY example, the phenylacetylglutamine (PAG) multiplet at 7.43 ppm was correlated with the whole spectra and peaks arising from the same

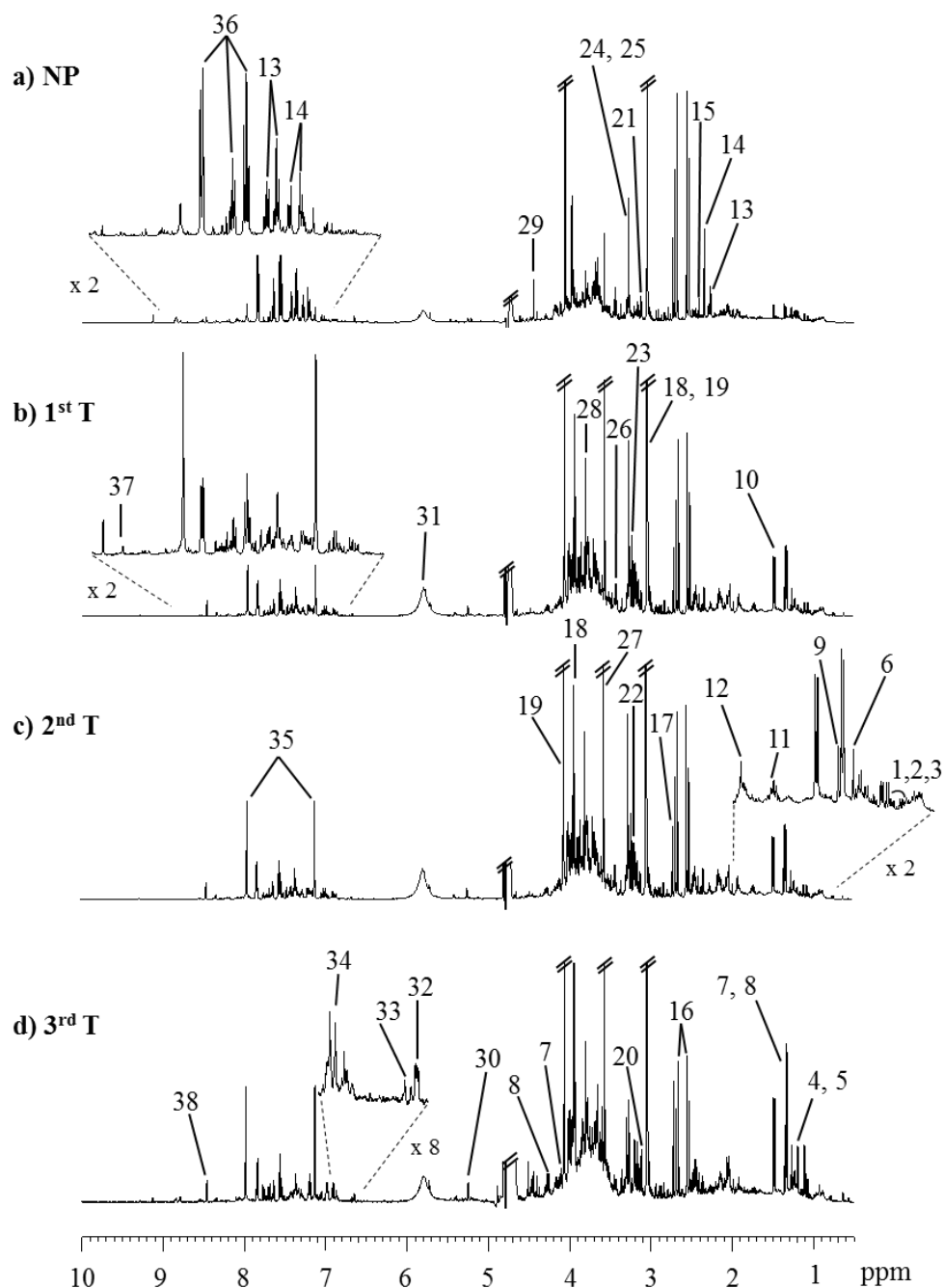


Figure 3.1: 500 MHz ^1H NMR urine spectra of a) NP and pregnant women at b) 1st T, c) 2nd T and d) 3rd T. Assignments (amino acids in 3-letter code): 1: Ile 2: Leu, 3: Val, 4: 3-hydroxybutyrate, 5: 3-aminoisobutyrate, 6: 3-hydroxyisovalerate, 7: lactate, 8: Thr, 9: 2-hydroxyisobutyrate, 10: Ala, 11: Lys, 12: acetate, 13: phenylacetylglutamine (PAG), 14: *p*-cresol sulphate (*p*-CS), 15: succinate, 16: citrate, 17: dimethylamine, 18: creatine, 19: creatinine, 20: malonate, 21: *cis*-aconitate, 22: choline, 23: carnitine, 24: betaine, 25: trimethylamine-*N*-oxide (TMAO), 26: taurine, 27: Gly, 28: guanidoacetate (GAA), 29: trigonelline, 30: glucose, 31: urea, 32: furoylglycine, 33: *N*-methyl-2-pyridone-5-carboxamide (2-Py), 34: Tyr, 35: His, 36: hippurate, 37: hypoxanthine, 38: formate. The rectangles guide the eye for some visible spectral changes. Remaining assignments are shown in Table 3.1.

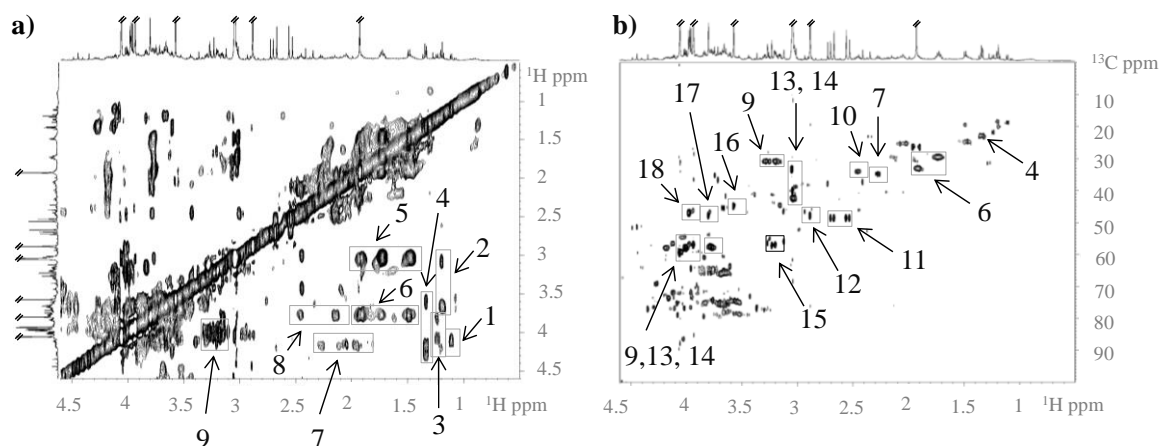


Figure 3.2: Expansions of the 2D NMR TOCSY and HSQC spectra of urine (respectively a and b), in the region from 0.5-4.6 ppm. Legend (amino acids in 3-letter codes): 1: 4-deoxyerythronic acid (4-DEA), 2: 3-aminoisobutyrate, 3: 4-deoxythreonic acid (4-DTA), 4: thr, 5: lys, 6: 2-ketoglutarate (2-KG), 7: phenylacetylglutamine (PAG), 8: gln, 9: his, 10: succinate, 11: citrate, 12: trimethylamine (TMA), 13: creatine, 14: creatinine, 15: choline, 16: gly, 17: guanidoacetate (GAA), 18: hippurate.

molecule (multiplet at 1.93 and 2.11, triplet at 2.27, doublet at 3.67, multiplet at 4.18, 7.36, and 7.43 ppm) showed high covariance and correlation ($r > |0.90|$), thus enabling the assignment of the complete spin system. The advantages of STOCSY over TOCSY is that it resolves the problem of overlapped resonances (occurring in 2D also) as well as allows the identification of the metabolically related compounds, due to their biological covariance. In this example, PAG was found to have high correlation and covariance with *p*-cresol sulphate (*p*-CS, indicated with * in figure), a known metabolically related compound, as both metabolites are produced from protein degradation by the gut microflora and resultant of the breakdown of tyrosine conjugated with glutamine and breakdown of phenylalanine and conjugated with sulphate, respectively for PAG and *p*-CS (Heinzmann *et al.*, 2012).

An example of a spiking experiment to assign betaine and TMAO singlets is shown in Figure 3.3b. In this figure, a urine spectrum is shown in black and, for the same sample, after the addition of betaine and TMAO standards, respectively in pink and green. The resulting spectra enable the assignment of these two adjacent singlets at 3.27 ppm and 3.28 ppm, respectively betaine and TMAO, as well as the second betaine singlet at 3.91 ppm. In spite of this being an ancient idea, spiking experiments are extremely helpful to identify singlets (which have no correlation through TOCSY or STOCSY) but also other low intensity peaks, which can be difficult to identify through TOCSY or HSQC.

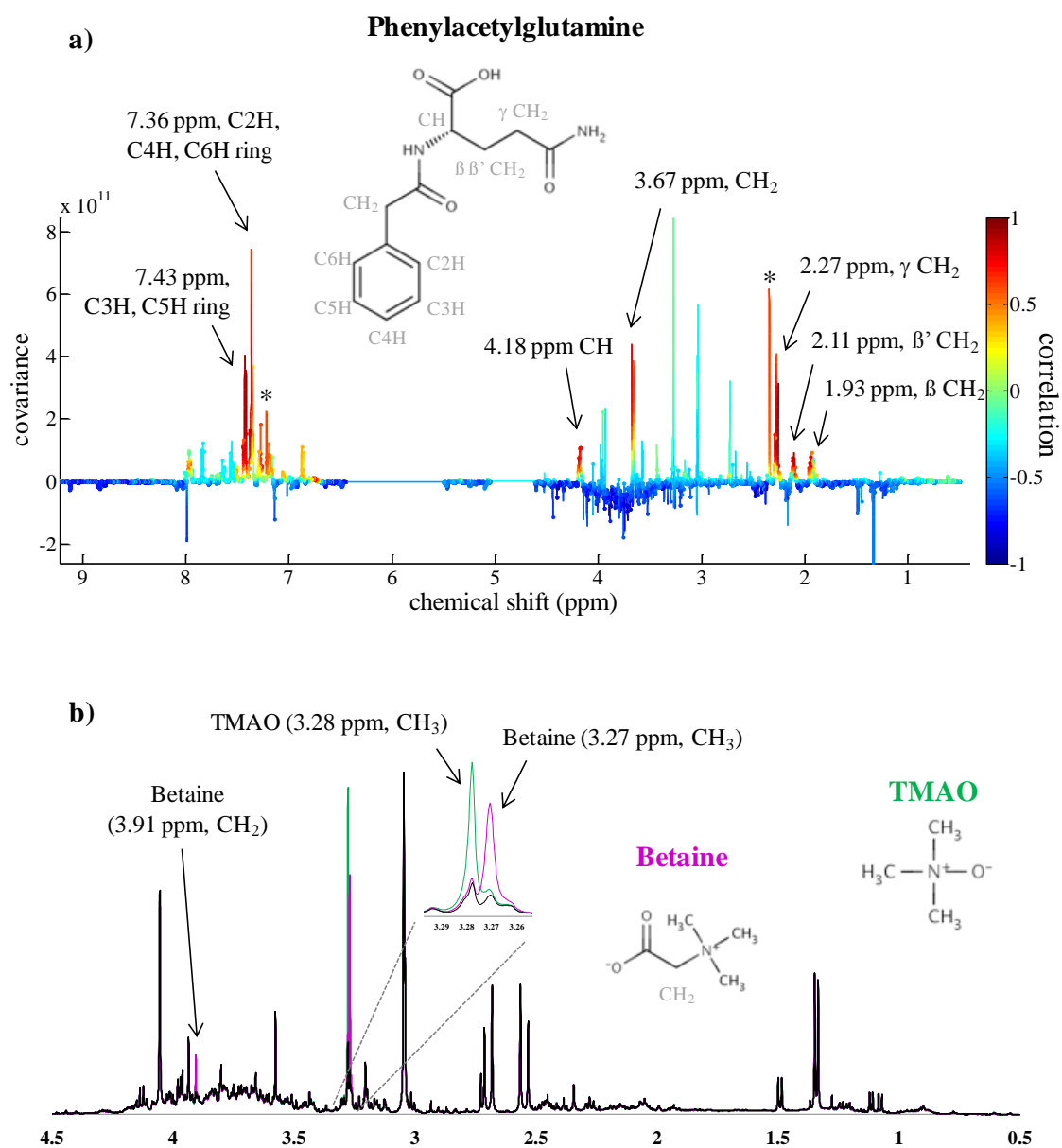


Figure 3.3: a) STOCSY for the assignment of phenylacetylglutamine (PAG) spin system and the metabolically related compound *p*-cresol sulphate (*p*-CS, peaks indicated with *) and b) example of spiking experiment for the assignment of betaine (pink) and trimethylamine-*N*-oxide (TMAO, green).

Overall, 64 metabolites were identified in maternal urine, including amino acids and derivatives, organic acids and sugars, all having been previously identified in urine (Rezzi *et al.*, 2007, Appiah-Amponsah *et al.*, 2009, Wishart *et al.*, 2009, Ellis *et al.*, 2012, Rivière *et al.*, 2012, Sachse *et al.*, 2012, Bouatra *et al.*, 2013). Peak identification was confirmed by comparison with databases, namely the human metabolome database (HMDB) (Wishart *et al.*, 2009, Bouatra *et al.*, 2013) and Bruker Biorecode 2.0.0 (courtesy

of Bruker BioSpin, Rheinsteten, Germany), as well as comparison with previous reports (Rezzi *et al.*, 2007, Appiah-Amponsah *et al.*, 2009, Ellis *et al.*, 2012, Rivière *et al.*, 2012, Sachse *et al.*, 2012). The complete list of assignments obtained is shown in Table 3.1 along with the corresponding chemical shifts, multiplicity, assignment and, in some cases, carbon chemical shift (measured in HSQC spectra). Metabolites confirmed by spiking are noted with ^a and those confirmed with STOCSY are noted with ^b in the same table.

Table 3.1: List of metabolite assignments in maternal urine ¹H NMR spectra, recorded in a 500 MHz spectrometer, pH 7.00±0.02. Legend: s: singlet, d: doublet, t: triplet, q: quartet, dd: doublet of doublets, m: multiplet. ^aMetabolites assigned through spiking and ^b through STOCSY.

Assignment	δ_H ppm (multiplicity, assignment/ δ_C ppm)
1,6-anhydroglucose	5.46 (br, C1H), 3.69 (m, C3H, C4H), 3.54 (m, C2H), 4.10 (dd, CH), 4.62 (dd, CH ₂), 3.76 (dd, CH ₂)
1-methyl-histidine ^a	3.07 (dd, β CH ₂); 3.16 dd, β' CH ₂); 3.72 (s, CH ₃); 3.96 (dd, α CH ₂); 7.05 (s, ring); 7.78 (s, ring)
2-hydroxyisobutyrate (2-HIBA)	1.36 (s, CH ₃)
2-ketoglutarate (2-KG)	3.45 (t, β CH ₂), 3.01 (t, γ CH ₂)
3-aminoisobutyricacid	1.19 (d, CH ₃ /17.87); 2.61 (m, α CH); 3.06 (dd, β CH ₂)
3-hydroxybutyrate (3-HBA)	1.2 (d, CH ₃); 2.31(m, CH ₂); 2.41 (m, CH ₂); 4.15 (m, CH)
3-hydroxyisovalericacid (3-HIVA)	1.27 (s, β CH ₃ /31.02); 2.37 (s, α CH ₂)
3-methyl-histidine ^a	3.28 (dd, β CH ₂ /28.05); 3.75 (s, NCH ₃ /35.18); 3.97(dd, α CH/56.52); 7.15 (s, C6H/126.33); 8.12 (s, C2H/140.81)
4-deoxyerythronic acid (4-DEA)	1.11 (d, γ CH ₃ /18.26); 4.08 (d, α CH/78.67); 4.10 (m, β CH/71.54)
4-deoxythreonic acid (4-DTA)	1.23 (d, γ CH ₃ /21.37); 3.84 (d, α CH/79.04); 4.12 (m, β CH/71.59)
4-hydroxyhippurate	3.95 (s, CH ₂); 6.98 (d, C3H, C5H ring); 7.76 (d, C4,2H, C6H ring)
4-hydroxyphenylacetate (4-HPA)	3.46 (s, CH ₂ /46.3); 6.86 (d, C3H, C5H ring/118.4); 7.17 (d, C2H, C6H ring/133.3)
Acetate	1.93 (β CH ₃)
Acetoacetate	3.46 (s, CH ₃); 2.29 (s, CH ₂)
Acetone	2.24 (s, CH ₃)
Acetyl-carnitine	2.15 (s, CH ₃), 2.5 (dd, α CH ₂), 2.61 (dd, α CH ₂), 3.19 (s, N(CH ₃) ₃), 3.61 (d, β CH ₂), 3.85 (dd, β CH ₂)
Alanine	1.49 (d, β CH ₃ /18.99); 3.78 (q, α CH)
Allantoin	5.39 (s, CH/66.19)
Ascorbate ^a	3.76 (m, CH ₂ (OH)); 4.01 (m, CH (OH)); 4.52 (d,C1H)
Betaine ^a	3.27 (s, CH ₃ /56.13); 3.91 (s, CH ₂ /69.11)
Carnitine	2.44 (dd, α CH ₂ /45.74; 3.23 (s, N(CH ₃) ₃ /72.80; 3.43 (m, γ CH ₂ /72.80; 4.57 (m, β CH ₂ /66.88)
Choline	3.20 (s, N(CH ₃) ₃ /56.52); 3.52 (m, NH/70.60); 4.07 (m, CH ₂ (OH))
cis-aconitate	3.12 (d, CH/46.13); 5.79 (t, CH ₂ /127.13)
Citrate	2.54 (d, α,β CH ₂ /48.17); 2.69 (d, α',β' CH ₂ /48.17)
Creatine	3.04 (s, NCH ₃ /39.66); 3.94 (s, NCH ₂ /56.55)
Creatinine	3.05 (s, NCH ₃ /32.86); 4.06 (s, NCH ₂ /59.05)

Assignment (<i>continues</i>)	δ_{H} ppm (multiplicity, assignment/ δ_{C} ppm)
Dimethylamine (DMA)	2.73 (s, CH ₃ /37.41)
Dimethylglycine (DMG)	2.93 (s, (CH ₃) ₂ /46.21); 3.72 (s, CH)
Formate	8.46 (s, CH/173.97)
Fumarate	6.53 (s, CH)
Furoylglycine ^b	3.93 (s, CH ₂); 6.65 (dd, C4H ring/114.98); 7.19(d, C3H ring); 7.70 (d, C5H ring)
Galactose	3.49 (dd, C4H), 3.64 (dd, C3H), 3.75 (m, C1H, C2H, CH ₂), 3.83 (m, C3H), 3.93 (d, C2H), 3.98 (d, C2H), 4.10 (t, C1H), 4.60 (d, CH ₂), 5.28 (d, C5H)
α -Glucose	3.23 (dd, C2H); 3.44 (m, C4H); 3.50 (t, C3H); 3.72 (dd, C6H') ; 3.90 (m, C6H); 4.65 (d, C1H)
β -Glucose	3.42 (t, C4H); 3.54 (dd, CH); 3.71 (t, C3H); 3.77 (dd, C6H); 3.84 (m, C5H) ; 5.25 (d, C1H/94.97)
Glutamine ^a	2.15 (m, β CH ₂ /29.24); 2.47 (m, γ CH ₂ /33.67); 3.79(t, α CH/57.41)
Glycine	3.57 (s, α CH ₂ /44.45)
Guanidoacetate (GAA)	3.80 (s, CH ₂ /47.44)
Hippurate	3.97 (d, CH ₂ /46.65); 7.56 (t, C4H, C6H ring/131.60); 7.64 (t, C3H, C5H ring/134.99); 7.83(d, C4H/129.97); 8.52 (br, NH)
Histidine	3.18 (dd, β CH ₂ /30.40); 3.28 (dd, β' CH ₂ /30.40);4.01(dd, α CH ₂ /57.64); 7.13 (s, C4H ring/120.05); 7.98 (s, C2H ring/138.74)
Hypoxanthine	8.20 (s, C2H ring); 8.22 (s, C8H)
Indoxyl sulphate (IS)	7.21 (dd, C8H/122.53), 7.28 (dd, C7H/125.01), 7.36 (s, C2H/119.08), 7.51 (d, C6H/115.04), 7.70 (d, C9H/120.57)
Isoleucine	0.94 (t, δ CH ₃); 1.01 (d, β CH ₂); 1.26 (m, γ CH ₂); 1.47(m, β' CH ₂); 1.98 (m, γ' CH ₂); 3.62 (d, α CH)
Lactate	1.34 (d, CH ₃ /22.43); 4.11 (q, CH/71.53)
Lactose	3.28 (dd, C2H); 3.55 (m, C'2H); 3.59 (dd, C2H); 3.66 (m, C'3H, C3H,C5H); 3.73 (m, C'6H, C'5H); 3.79 (m, C6H); 3.86 (m, C6H, C3H); 3.94 (m, C6H, C'4H, C4H); 4.46 (d, C'1H/ 105.8); 5.25 (d, C1H)
Leucine	0.96 (t, γ CH ₃); 1.7 (m, CH ₂); 3.73 (t, α CH)
Lysine	1.48 (m, γ CH ₂ /24.28); 1.73 (m, δ CH ₂ /29.06); 1.92 (m, β CH ₂ /32.60); 3.03 (t, ϵ CH ₂ /42.02); 3.77 (t, α CH/57.28)
Malonate	3.11 (s, CH ₂ /48.7)
<i>N</i> -methyl-2-pyridone-5-carboxamide (2PY) ^b	3.62 (s, CH ₃); 6.66 (d, C3H ring/120.85); 7.97 (dd, C4H ring); 8.33 (d, C6H ring/145.46)
<i>N</i> -methylnicotinamide (NMND)	4.48 (s, NCH ₃ /51.30); 8.18 (m, C5H ring); 8.90 (d, C4H ring); 8.97 (d, C6H ring); 9.29 (s, C2H ring)
<i>p</i> -cresol sulphate (<i>p</i> -CS) ^b	2.35 (s, CH ₃); 7.21 (d, C2H, C6H ring/124.12); 7.29 (C3H, C5H ring/125.04)
Phenylacetylglutamine (PAG) ^b	1.93 (m, β CH ₂) ; 2.11 (m, β' CH ₂); 2.27 (t, γ CH ₂ /34.44); 3.67 (d, CH ₂); 4.18 (m, α CH); 7.36 (m, C2H, C4H, C6H ring/132.01); 7.43 (m, C3H, C5H ring/131.84)
Pyruvate ^a	2.39 (s, CH ₃)
<i>Scyllo</i> -inositol ^a	3.37 (s, CH)
Succinate	2.41 (s, CH ₂ /36.85)
Sucrose	3.48 (t, C4H), 3.56 (dd, C2H), 3.63 (s, C1'H ₂), 3.77 (t, C3H), 3.83 (dd, CH ₂ , C'6H ₂), 3.85 (m, C5H), 3.89 (m, C'5H,), 4.06 (t, C'4H), 4.22 (d, C'3H), 5.41 (d, C1H)

Assignment (<i>continues</i>)	δ_{H} ppm (multiplicity, assignment/ δ_{C} ppm)
Tartrate ^a	4.35 (s, CH(OH))
Taurine ^a	3.26 (t, CH ₂ SO ₃); 3.43 (t, NCH ₂)
Threonine	1.33 (d, CH ₃ /22.44); 3.61 (d, β CH/63.36); 4.26 (dd, α CH/68.90)
Trigonelline	4.44 (s, CH ₃ /50.98); 8.09 (t, C3H ring); 8.84 (br, C2H, C4H ring); 9.12 (s, C6H ring/148.73)
Trimethylamine (TMA)	3.89 (s, CH ₃ /47.40)
Trimethylamine- <i>N</i> -oxide ^a (TMAO)	3.28 (s, CH ₃ /62.31)
Tyrosine	3.06 (dd); 3.21 (dd); 3.95 (dd); 6.91 (d, C3H, C5H ring/118.83); 7.20 (d, C2H, C6H ring/124.15)
Urea	5.79 (br s, NH ₂)
Valine	0.99 (d, γ CH ₃); 1.04 (d, γ' CH ₃); 2.27 (m, β CH); 3.61 (d, α CH)
Xylose ^a	3.23 (dd, C3H), 3.33 (dd, C6H), 3.42 (t, C4H), 3.53 (dd, C3H), 3.63 (m, C6H, C5H, C4H), 3.93 (dd, C6H), 4.59 (d, C2H), 5.21 (d, C2H)

3.2 Impact of non-fasting conditions on urine metabolomic studies

In this work, due to limitations in controlling/restricting pregnant women diet in the daily hospital setting, all samples were collected in non-fasting conditions, during the morning period and after non-uniformised breakfast. This apparent constraint can become, in fact, one strong point of the study since developing multivariate models which are independent of confounding effects (arising from dietary and lifestyle habits), enables clinical applications to be envisaged, in situations where the ideal fasting condition (or standardized meals) are not possible or easy e.g. biofluid analysis at check-in at the clinic for immediate patient evaluation.

Many previous metabolomics studies have evaluated the effect of diet in urine composition, usually by subjecting participants to controlled diets and collecting samples after predefined times, these interventions often being lengthy (e.g. weeks). In this context, the effect of several diets has been investigated, showing relevant changes in urine metabolomic profiles after consumption of specific foods, namely meat or animal protein (Stella *et al.*, 2006, Bertram *et al.*, 2007, Heinzmann *et al.*, 2012), vegetarian diets (Stella *et al.*, 2006), soy-rich diets (Solanky *et al.*, 2005), tea (Wang *et al.*, 2005, van Velzen *et al.*, 2009), chocolate (Rezzi *et al.*, 2007, Martin *et al.*, 2009), fruits (Heinzmann *et al.*, 2012), wine and grapes (Heinzmann *et al.*, 2012), lingoberries (Lehtonen *et al.*, 2013). Also, the effect of long-term dietary habits have been explored (Zuppi *et al.*, 1998, Lenz *et al.*, 2004, Holmes *et al.*, 2008), a thorough review of the metabolic effects of nutritional

interventions having been recently published (Llorach *et al.*, 2012). Moreover, studies comparing urine samples collected fasting and non-fasting, after standardized breakfast, have also been pursued using a mass spectrometry approach (Favé *et al.*, 2011, Lloyd *et al.*, 2011). These studies showed that urinary changes between samples collected at fasting and postprandial were robust and reproducible within individuals suggesting a 2-4 h postprandial window for urine collection after a standardized breakfast (Favé *et al.*, 2011). Furthermore, the effect of time of day for urine collection, morning *vs.* afternoon, has also been addressed (Saude *et al.*, 2007, Slupsky *et al.*, 2007) showing the susceptibility of urine composition during the day due to diet or circadian rhythms.

3.2.1 Effect of non-fasting on urine composition

In the present work, in order to evaluate the impact of non-fasting in the urinary metabolome in a female population, samples were collected for a group of NP in overnight fasting conditions and then approximately 2 hours after random breakfast, conditions in which collection occurred for all pregnant women enrolled in this study. The score plots of both PCA and PLS-DA showed higher dispersion of samples collected non-fasting, compared to those collected fasting (Figure 3.4). However, in spite of the apparent separation seen in the PLS-DA score plot, the model was found to be weak, as given by the Q^2 value of 0.295, and model robustness assessed by MCCV yielded a classification rate of 66%, sensitivity 65% and specificity 80% and approximately 40% of true models having $Q^2 \leq -1$. Nonetheless, the analysis of the loading plot unveiled differences in some metabolites, these being further evaluated by spectral integration and univariate comparison.

Relevant and significant variations included decreased creatinine and hypoxanthine (effect size higher than -1) and increased furoylglycine (effect size 1.93), four unassigned spin systems/spectral regions (singlet at 2.48 ppm, singlet at 4.40 ppm, doublet at 8.45 ppm and region 6.47-6.57 ppm) in non-fasting, these being listed in Table 3.2 along with the corresponding effect sizes and *p*-values. Whithin these metabolites, only creatinine and furoylglycine had been previously linked to dietary interventions (Kumps *et al.*, 2002, Solanky *et al.*, 2005, Wang *et al.*, 2005, Stella *et al.*, 2006) but not hypoxanthine.

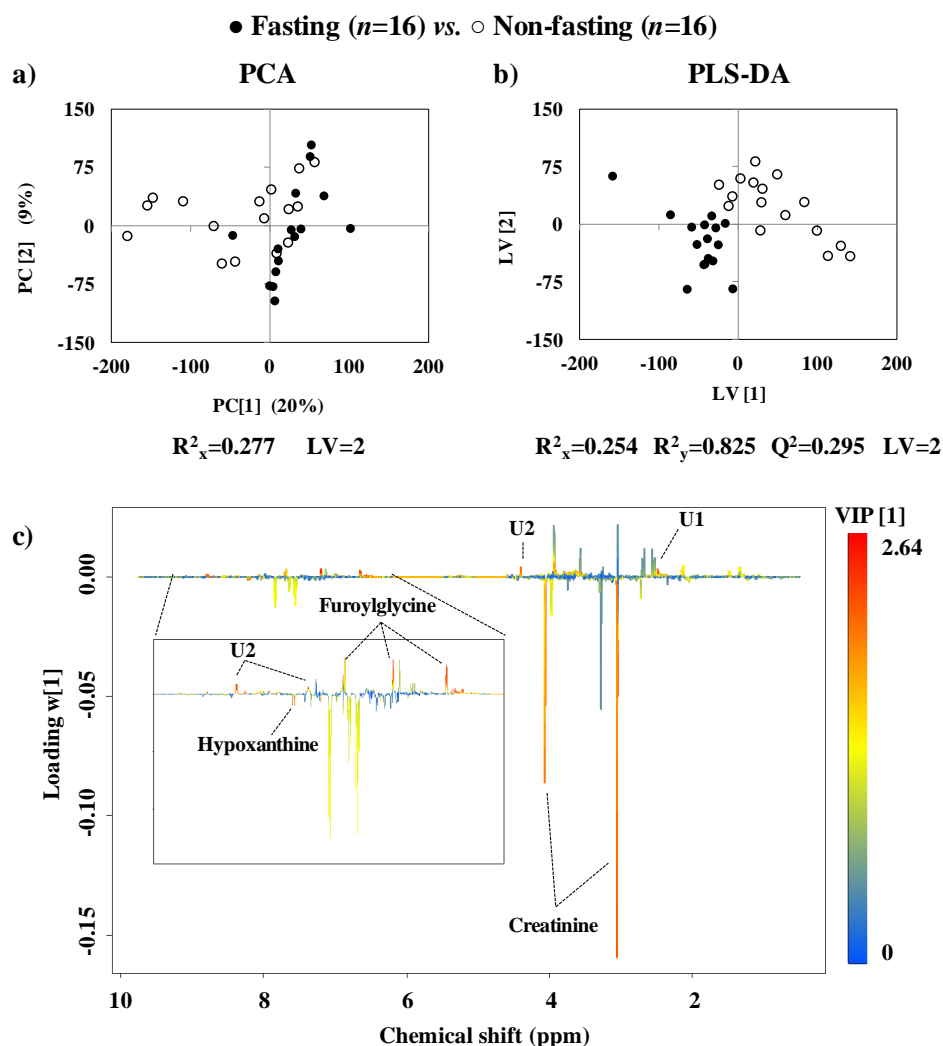


Figure 3.4: a) PCA and b) PLS-DA scores scatter plots of ^1H NMR spectra of urine from NP women collected under fasting (●) and non-fasting (○) conditions and c) corresponding PLS-DA loading weights $w[1]$ colored by VIP [1]. U_i : unassigned resonance i in order of appearance in Table 3.2, R^2_x : variation explained by the **X** matrix R^2_y : variation explained by the **Y** matrix, LV: latent variable.

3.2.2 Proposed metabolic interpretation of the urinary changes under non-fasting

Urinary creatinine has been previously found to vary with diet, namely increased after chamomile tea (Wang *et al.*, 2005) and meat (Stella *et al.*, 2006) and decreased after soy-rich diets consumption (Solanky *et al.*, 2005). Furoylglycine is formed by the conjugation of glycine with 2-furoic acid, the excretion of this compound having been related with diet, namely consumption of chocolate and heated juices (Kumps *et al.*, 2002).

Increased urinary hypoxanthine have been associated with impaired ATP metabolism and in ischemia (Turgan *et al.*, 1999) and after physical exercise (Enea *et al.*, 2010), but not in relation to diet. Regarding unassigned resonances, a relevant variation was found for U1 (effect size 2.28), along with U2 and U3 (effect size higher than 1.60), and the unassigned region 6.47-6.57 (effect size 1.91).

Table 3.2: List of urinary changes noted, for the group of NP women, under non-fasting conditions compared to overnight fasting conditions. ^a Chemical shifts correspond to signals used for integration; values in square brackets correspond to peaks correlated through STOCYSY; s: singlet, d: doublet, dd: doublet of doublets; ^b Ui: unassigned resonances number in order of appearance. ^c Significance level 95% (p -value <0.05). ^d Variation shown as effect size.

Chemical shift (multiplicity) ^a	Assignment ^b	Variation (effect size, p -value) ^{c,d}
2.48 (s)	U1	↑ (2.28, 9.5×10^{-6})
4.06 (s)	Creatinine	↓ (-1.10, 3.9×10^{-3})
4.40 (s) [8.05, 8.54, 8.78]	U2	↑ (1.60, 8.3×10^{-3})
6.65 (dd)	Furoylglycine	↑ (1.93, 3.0×10^{-5})
8.20 (s), 8.22 (s)	Hypoxanthine	↓ (-1.35, 1.5×10^{-3})
8.45 (d)	U3	↑ (1.64, 7.4×10^{-3})
6.47-6.57		↑ (1.91, 5.5×10^{-5})

Therefore, it may be concluded that these studies could benefit from sample collection under fasting conditions, specifically towards controlling the contribution of diet as a confounding effect and in reducing sample dispersion. Nevertheless, the overall impact of non-fasting in multivariate models was found to be weak (supported by MCCV results). On the other hand, univariate analysis of individual resonances' intensities confirmed the statistical (p -values 10^{-3} - 10^{-6}) and biological significance (effect size higher than |1|) of 7 spin systems/spectral regions, meaning that a possible contribution of the non-fasting condition must be considered when interpreting variations of these metabolites.

3.3 Following healthy pregnancy metabolism by NMR metabolic profiling of maternal urine

3.3.1 Effect of healthy pregnancy on urine metabolite profile

In this study, comparison of urine from healthy NP and pregnant women in their 1st, 2nd and 3rd trimesters was carried out in order to characterize the time course metabolic

changes of healthy pregnancies, as viewed through the metabolite composition of maternal urine. This multipoint collection study unveils metabolite changes describing the expected healthy pregnancy progression so that, in future studies, a deviating trajectory in pathological pregnancies may be detected as a means of diagnostic.

Typical ^1H NMR spectra of urine collected for non-pregnant (NP) women and pregnant women in their 1st, 2nd and 3rd trimesters (T) of pregnancy are shown in Figure 3.1 along with some relevant assignments. Indeed, some relevant changes can be visually inferred such is the case of isoleucine (\uparrow 3rd T), leucine (\uparrow 3rd T), 4-deoxyerythronic acid (4-DEA, \uparrow 1st and 3rd T), 3-hydroxybutyrate (3-HBA, \uparrow 3rd T), alanine (\uparrow 1st, 2nd and 3rd T), guanidoacetate (GAA, \uparrow 1st T), histidine (\uparrow 1st T) and tyrosine (\uparrow 1st T), indicated by rectangles in the same figure.

Urinary metabolic changes characteristic of each pregnancy trimester

In order to confirm the visually observed changes, mentioned in the above paragraph, and identify other relevant changes in the ^1H NMR spectra, as a function of gestation time, MVA was carried out considering each pair of sample groups, i.e. NP vs. 1st T, 1st T vs. 2nd T and 2nd T vs. 3rd T. Model robustness and predictive power was assessed by MCCV (quality parameters shown in Annex III, Table A-III.1 and Figure A-III.1) and the PLS-DA model scores and loadings plots obtained are shown in Figure 3.5.

The PLS-DA scores plot corresponding to NP vs. 1st T samples (Figure 3.5a), shows a clear group separation, with high Q^2 (0.727) expressing its reliability. MCCV revealed 90% sensitivity and specificity, 88% classification rate and high median Q^2 value (0.724), reflecting the good predictive power of this model. These results highlight that the major metabolic adaptations occurring in the 1st T of pregnancy, compared to the NP state, have a significant impact on urine composition. This high impact in 1st T urine was somehow expected based on previous reports, which had shown significant variations of selected amino acids in urine of pregnant and non-pregnant women (Miller *et al.*, 1954, Hytten, 1973, Cho and Cha, 2005, Creasy *et al.*, 2009).

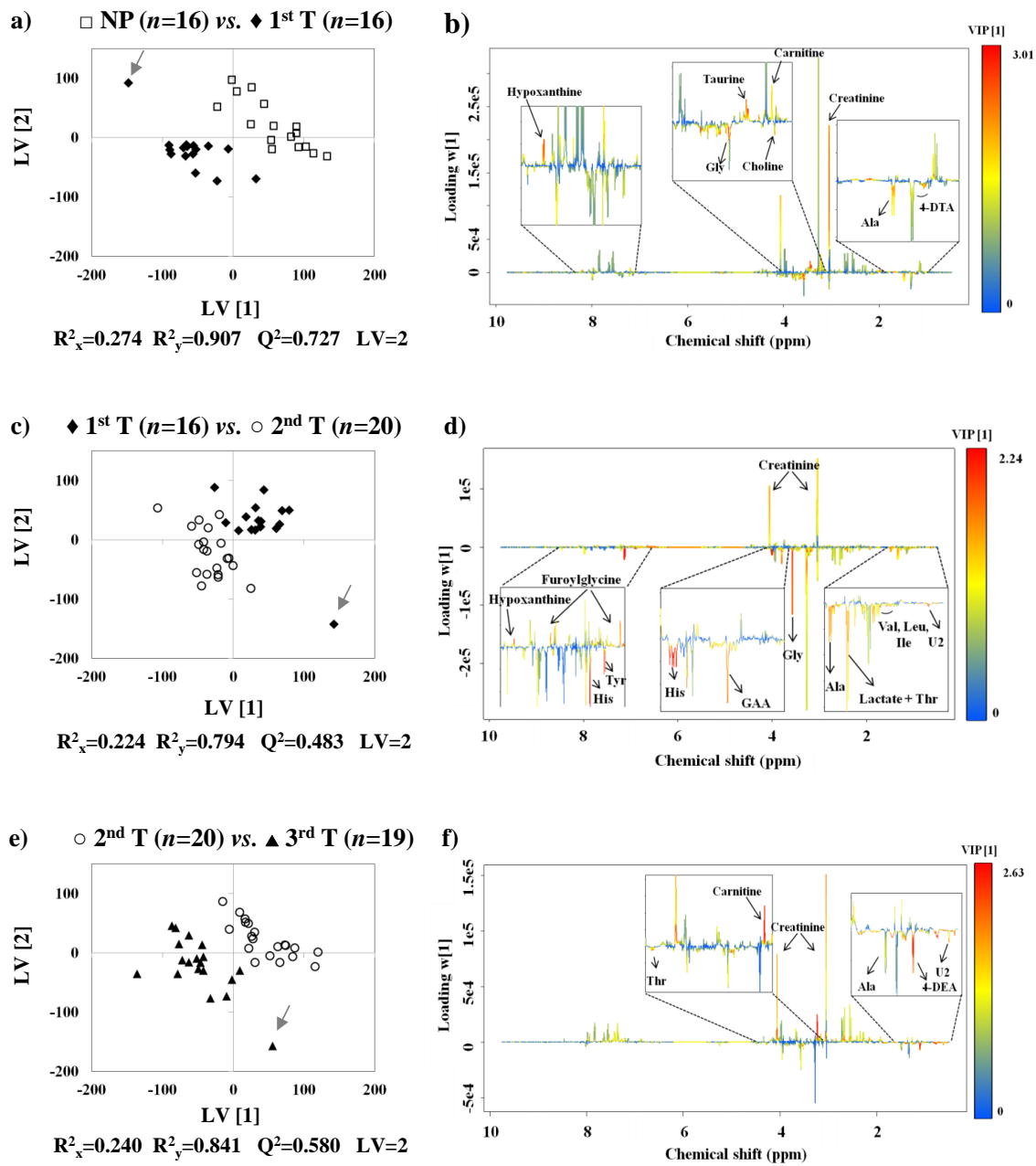


Figure 3.5: PLS-DA scores and loading plots, colored by the VIP value, for a,b) NP vs. 1st T, c,d) 1st vs. 2nd T and e,f) 2nd vs. 3rd T. Ui: unassigned spin systems defined in Table 3.3, three-letter code for amino acids, GAA: guanidoacetate, 4-DTA: 4-deoxythreonic acid, 4-DEA: 4-deoxyerythronic acid. R^2_X : variation explained by the **X** matrix R^2_Y : variation explained by the **Y** matrix, LV: latent variable. Arrows indicates outlier samples.

The PLS-DA score plots of 1st vs. 2nd T and 2nd vs. 3rd T show a clear separation between pregnancy trimesters (Figure 3.5) supported by Q^2 values of 0.483 and 0.580, respectively, although to with a smaller magnitude than the NP vs. 1st T. MCCV, however, of the 1st vs. 2nd T model showed lower classification rate, 70%, lower sensitivity, 75%,

and especially lower specificity, 64% (higher rate of false positives), along with a low median Q^2 value (0.324). The lower predictive power of this model, in the particular conditions of this study, is explained by the closeness of gestational ages (medians of 12 and 17 g.w., respectively for 1st and 2nd trimesters (Table 2.2, Experimental Section) and also reflects that metabolic changes between 1st and 2nd trimesters are not as marked as in the previous pregnancy stage. Regarding the 2nd vs. 3rd T model, MCCV showed improved classification rate (80%), specificity and sensitivity (82% and 86%, respectively), with median Q^2 of 0.416. Regarding to the two samples noted in Figure 3.5 (NP vs. 1st T, 1st vs. 2nd T and 2nd vs. 3rd T), which represent two samples collected at 1st and 3rd T clearly distant from the bulk of the 1st and 3rd T groups. The visual inspection of the corresponding spectra of these outlier samples revealed high concentrations of hippurate and TMAO, both metabolites possibly relating to diet (Llorach *et al.*, 2012).

The analysis of the corresponding loading weights plots for each model enabled the identification of the main compounds varying between the four groups. Relevant metabolites were integrated and tested for statistical (p -value<0.05) and biological (effect size) significance. Overall, 21 metabolites were found to vary significantly at some stage during pregnancy, their variations (described by direction, effect size and p -value) being listed in Table 3.3. Out of this 21-metabolite signature, 9 were in agreement with literature (namely leucine, isoleucine, alanine, carnitine, taurine, glycine, threonine, tyrosine and histidine) while 8 metabolites were registered to vary in connection to pregnancy for the first time (namely 4-DEA, 4-deoxythreonic acid (4-DTA), 3-HBA, choline, GAA, creatinine, lactate and furoylglycine) these being noted with * in Table 3.3, the remaining 4 metabolites being still unassigned.

The results obtained showed that the urinary profile of the 1st T was characterized by the increased excretion of alanine, glycine, histidine, tyrosine, 4-DEA, GAA and four unassigned metabolites (singlet at 0.55, 0.63 and 2.15 ppm and a doublet at 1.18 ppm) along with decreased creatinine and furoylglycine. Out of these, only U2 (singlet at 0.63 ppm) and alanine increased significantly throughout pregnancy (the variation of alanine being expected based on literature (Hyttén, 1973, Creasy *et al.*, 2009)). Regarding the 2nd T, urine exhibits increased alanine, U2, 4-DTA and glycine along with decreased taurine

Table 3.3: List of metabolites/resonances changes in urine from the NP state to 1st, 2nd and 3rd trimesters. ^aChemical shifts shown correspond to signals used for integration; value in square brackets for correspond to correlated peak seen through TOCSY or STOCYSY; s: singlet, d: doublet, t: triplet, q: quartet, dd: doublet of doublets, m: multiplet. ^bSignificance level: 95% (p -value<0.05). ^cMagnitude of variations shown as effect size. Ui: unassigned resonance for compound i. ^dPeak overlapped with other resonances. ^ePutatively assigned as acetyl-carnitine. *Metabolites observed to vary in maternal urine, for the first time throughout pregnancy. 4-DEA: 4-deoxyerythronic acid, 3-HBA: 3-hydroxybutyrate, 4-DTA: 4-deoxythreonic acid, GAA: guanidoacetate.

δ_H /ppm and multiplicity	Assignment	Variation (effect size ^c , p -value ^b)		
		NP vs. 1 st T	1 st T vs. 2 nd T	2 nd T vs. 3 rd T
0.55 s [0.78 m]	U1	$\uparrow(2.7, 3.1 \times 10^{-6})$	—	$\uparrow(2.47, 1.4 \times 10^{-7})$
0.63 s	U2	$\uparrow(3.3, 1.8 \times 10^{-6})$	$\uparrow(0.9, 1.1 \times 10^{-2})$	$\uparrow(2.55, 3.9 \times 10^{-7})$
0.96 t	Leucine	—	—	$\uparrow(0.92, 3.1 \times 10^{-3})$
1.01 d	Isoleucine	—	—	$\uparrow(2.07, 1.9 \times 10^{-6})$
1.11 d	4-DEA*	$\uparrow(0.78, 4.2 \times 10^{-2})$	—	$\uparrow(1.61, 1.3 \times 10^{-4})$
1.18 d [1.45, 3.82]	U3	$\uparrow(0.82, 2.7 \times 10^{-2})$	$\downarrow(-0.9, 1.4 \times 10^{-2})$	—
1.20 d	3-HBA*	—	—	$\uparrow(1.57, 6.1 \times 10^{-5})$
1.23 d	4-DTA*	—	$\uparrow(0.92, 1.7 \times 10^{-2})$	$\uparrow(0.86, 2.6 \times 10^{-2})$
1.49 d	Alanine	$\uparrow(1.02, 1.3 \times 10^{-2})$	$\uparrow(0.88, 3.4 \times 10^{-2})$	$\uparrow(0.63, 4.8 \times 10^{-2})$
2.15 s	U4 ^{d,e}	$\uparrow(0.84, 2.5 \times 10^{-2})$	—	—
3.20 s	Choline*	$\uparrow(0.88, 3.9 \times 10^{-2})$	—	—
3.23 s	Carnitine	—	—	$\downarrow(-1.76, 1.7 \times 10^{-5})$
3.43 t	Taurine	—	$\downarrow(-1.35, 1.3 \times 10^{-3})$	—
3.57 s	Glycine	$\uparrow(1.45, 1.6 \times 10^{-4})$	$\uparrow(0.68, 3.9 \times 10^{-2})$	—
3.80 s	GAA*	$\uparrow(1.19, 1.1 \times 10^{-3})$	—	—
4.06 s	Creatinine*	$\downarrow(-1.43, 3.3 \times 10^{-4})$	—	$\downarrow(-0.76, 3.5 \times 10^{-2})$
4.11 q	Lactate*	—	—	$\uparrow(0.37, 2.5 \times 10^{-2})$
4.26 m	Threonine	—	—	$\uparrow(1.22, 9.1 \times 10^{-4})$
6.65 dd	Furoylglycine*	$\downarrow(-1.92, 5.9 \times 10^{-5})$	—	$\uparrow(0.96, 1.9 \times 10^{-3})$
6.91 d	Tyrosine	$\uparrow(1.11, 4.4 \times 10^{-3})$	—	—
7.13 s	Histidine	$\uparrow(1.12, 1.1 \times 10^{-2})$	—	—

and U3 (doublet at 1.18 ppm). Within these changes that of alanine, glycine and taurine are consistent with previous literature reports (Hyttén, 1973, Cho and Cha, 2005, Creasy *et al.*, 2009) whilst that of 4-DTA was here found for the first time to our knowledge. Finally, third trimester urine was characterized by an increase in alanine, leucine, isoleucine, threonine, 4-DEA, 4-DTA, 3-HBA, U1 (singlet at 0.55 ppm), U2 and lactate with decreased creatinine and carnitine. Out of these changes, the increase of alanine, leucine, isoleucine and threonine were consistent with previous findings (Miller *et al.*, 1954, Hyttén, 1973, Creasy *et al.*, 2009), whilst creatinine, 4-DEA, 4-DTA and 3-HBA were here found to vary for the first time to our knowledge. Regarding the unassigned resonances

here found to vary across pregnancy, U1 and U2 (singlets at 0.55 and 0.63 ppm) possibly arise from bile acids as these type of compounds typically resonate in the 0.5-1.0 region (Trump *et al.*, 2006, Duarte *et al.*, 2009), U3 (doublet at 1.18 ppm) and U4 (singlet at 2.15 ppm) possibly belonging to acetyl-carnitine although definitive assignment was not possible due to the low intensity of the peak in this set of spectra and overlapping with other resonances.

Previous reports have indicated that a) the excretion of alanine, glycine, histidine, serine and threonine increases during pregnancy; b) the excretion of cysteine, leucine, lysine, phenylalanine, taurine, tyrosine and valine increases early on and decreases later in gestation and that c) the excretion of arginine, asparagine, glutamate, isoleucine, methionine and ornithine is not expected to change significantly (Hyttén, 1973, Creasy *et al.*, 2009). In this work, alanine, glycine, histidine and threonine were confirmed to increase in urine; changes in excreted taurine and tyrosine were confirmed, while leucine varied differently than expected (increased in 3rd T). Out of the amino acids not expected to change significantly, isoleucine presented a marked increase in the 3rd trimester, consistently however with a much earlier report (Miller *et al.*, 1954). In relation to the remaining amino acids, arginine, asparagine, cysteine, glutamate, methionine, phenylalanine and ornithine were not detected (probably due to their too low concentrations for NMR observation) whereas no significant changes could be measured for lysine, serine and valine. Moreover, carnitine has been reported to increase in the 1st T and decrease in the 3rd T (Cho and Cha, 2005), only the later trend having been found significant in this study.

To facilitate visual inspection of the variations, metabolites' integrals were plotted in a qualitative color-coded heatmap (Figure 3.6). In this plot, each line represents one subject and each column corresponds to a metabolite set of integrals, normalized to unity in order to enable the comparison of variations. This means that the maximum integral value is colored in dark red, the minimum value is shown in dark blue and the intermediate integrals are shown accordingly to the gradient of color. Inspection of this figure firstly reveals the significant extent of inter-subject variability. Still, and in spite of some dispersion noted in metabolite intensities, some of the above mentioned variations are clearly visible, namely the increases of U1 (0.55 ppm, singlet) and U2 (0.63 ppm, singlet),

amino acids alanine, glycine, histidine and tyrosine, choline, 4-DEA, 3-HBA and decreases for creatinine, carnitine and furoylglycine.

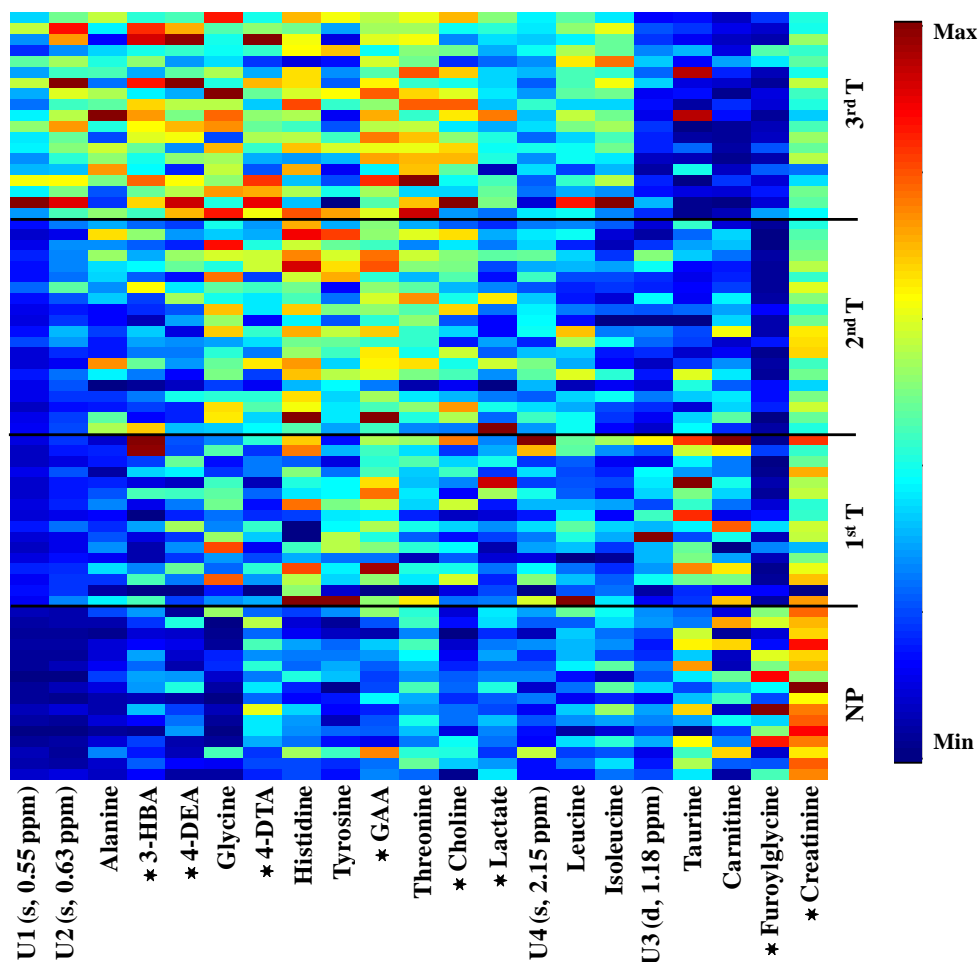


Figure 3.6: Heatmap of normalized integrals of the 21 metabolites varying across pregnancy. Lines and columns represent, respectively, subjects and integrals. Integrals are shown in a color scale, with dark blue representing the minimum integrals value and dark red the maximum integral value. Metabolites are ordered, from left to right, from those exhibiting marked increases to those exhibiting marked decreases across pregnancy (* denotes metabolites found to vary across pregnancy for the first time).

In order to verify that indeed the metabolic trajectory throughout pregnancy can be suitable described by this 21-metabolite profile, MVA was repeated using this selection of integrals alone. The resulting two-class models had a significant improvement in terms of CR, sensitivity and specificity determined by MCCV (NP vs. 1st T: CR 98%, sensitivity 98%, specificity 96%; 1st T vs. 2nd T: CR78%, sensitivity: 89%, specificity 68%; 2nd T vs. 3rd T: CR 99%, sensitivity 96%, specificity 100%, also show in Annex III, Table A-III.1

and Figure A-III.2) with the exception of the 1st vs. 2nd T model. Moreover, PCA and PLS-DA of the four classes using this set of 21 metabolites unveiled an overall metabolic trajectory of healthy pregnancies significantly improved compared to the whole ¹H NMR spectra (Figure 3.7b,c compared to a,b). The curved directional path in Figure 3.7d) indicates the trajectory followed throughout pregnancy and the sample indicated with a short arrow corresponds to one collected relatively late in pregnancy (39 g.w.), thus seemingly continuing the samples trend towards pregnancy term. This sample was shown to have high contents of choline and U1 (possibly a bile acid), these components revealing to be highly determinant at the end of pregnancy.

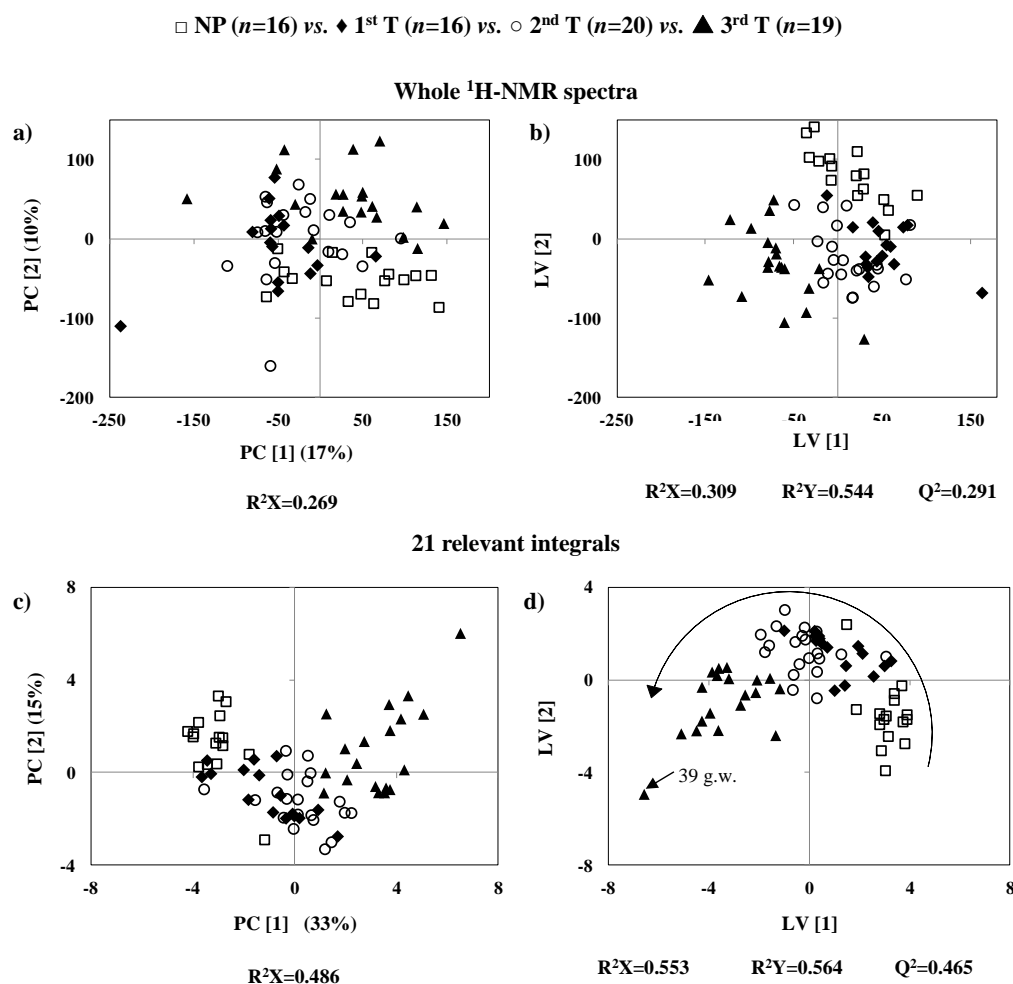


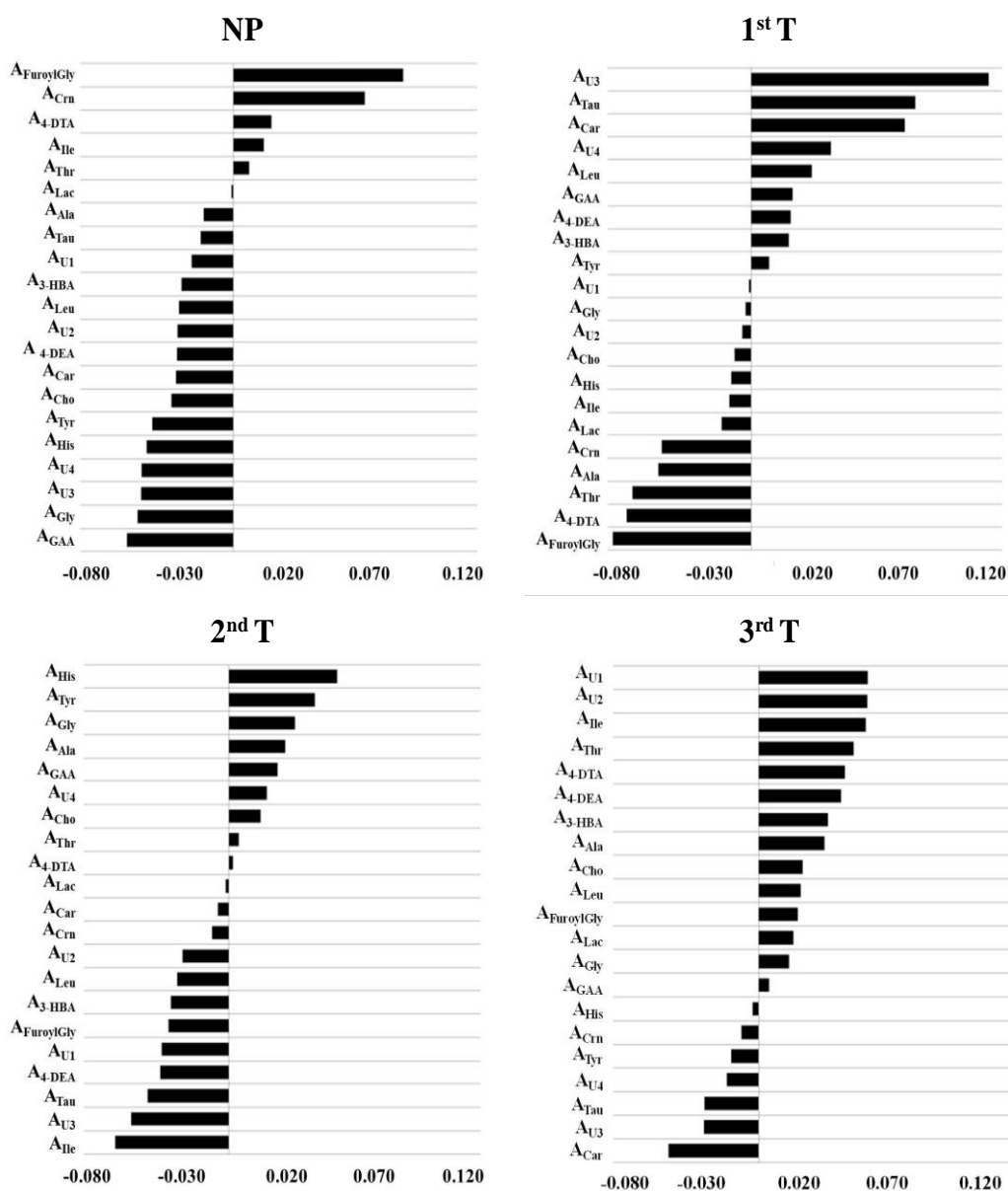
Figure 3.7: a,b) PCA and PLS-DA score plots obtained using the whole ¹H NMR spectra of urine and c,d) PCA and PLS-DA score plots obtained using the selection of 21 relevant integrals of NP, 1st T, 2nd T and 3rd T pregnant women. The arrow in d) indicates the trajectory across pregnancy. R²_x: variation explained by the X matrix R²_y: variation explained by the Y matrix, LV: latent variable.

Based on the PLS-DA obtained using the set of 21 metabolites (shown in Figure 3.7d), each subject group could then be described by a discrimination function of the type $y = c + b_1x_1 + \dots + b_nx_n$, with $n=21$, where each coefficient b_i gives the contribution of variable i for what can be considered a profile signature for each trimester of a healthy pregnancy. These functions describe the weight-distribution of the 21-metabolite integrals characteristic of NP and each pregnancy stage. The relative contribution of each variable, in this case each integral and corresponding equation for each trimester are shown in Figure 3.8. The usefulness of these equations is that they provide class membership for a given new sample. This means that by calculating each of the variables, in this case the area of each peak (integral), and multiplying that area by the corresponding b coefficient, the resulting linear combination yields a y value which converges towards 0, if the sample does not belong to that class, in this case that trimester, or 1 if the sample belongs to that trimester. It is envisaged that these function can be useful in future profile-matching of the urine of new subjects (in a given trimester) to confirm the healthy course of pregnancy or detect deviant trajectories.

Elucidation of metabolic relationships using STOCSY

Further metabolic characterization of each subject group was attempted through STOCSY, in order to identify eventual metabolically related compounds. Correlations between all 21 metabolites were calculated for each pregnancy trimester along with the corresponding p -value. A threshold of $r > 0.6$ and $p\text{-value} < 0.05$ was applied and correlations were confirmed by scatter plotting each pairs of correlated integrals (not shown) to detect spurious correlations.

Meaningful correlations were found for the 1st T group, namely between U1/U2 (r 0.70), isoleucine/leucine (r 0.68), isoleucine/U4 (r 0.73) and isoleucine/carnitine (r 0.72), 3-HBA/U4 (r 0.93) and 3-HBA/carnitine (r 0.74), alanine/GAA (r 0.61), 4-DTA/carnitine (r 0.74). For the 2nd trimester groups, relevant correlations were found for U1/U2 (r 0.88), isoleucine/leucine (r 0.63). The 3rd trimester group was characterized with correlations between U1/U2 (r 0.76), U2/3-HBA (r 0.74), 4-DEA/3-HBA (r 0.73), 4-DTA/3-HBA (r 0.64) and isoleucine/leucine (r 0.85). The above correlations enables the proposal of putative metabolic relationships between branched chain amino acids (BCAA) metabolism,



$$Y_{pr}[NP] = 0.225 - 0.022A_{U1} - 0.029A_{U2} - 0.028A_{Leu} + 0.016A_{Ile} - 0.029A_{4-DEA} - 0.048A_{U4} - 0.027A_{3-HBA} + 0.020A_{4-DTA} - 0.015A_{Ala} - 0.048A_{U4} - 0.032A_{Cho} - 0.030A_{Crn} - 0.017A_{Tau} - 0.049A_{Gly} + 0.055A_{GAA} + 0.069A_{Cm} - 0.001A_{Lac} + 0.009A_{Thr} + 0.089A_{FuroylGly} - 0.042A_{Tyr} - 0.045A_{His}$$

$$Y_{pr}[1^{st}T] = 0.225 - 0.001A_{U1} - 0.005A_{U2} + 0.034A_{Leu} - 0.012A_{Ile} + 0.022A_{4-DEA} + 0.13A_{U3} + 0.021A_{3-HBA} - 0.070A_{4-DTA} - 0.052A_{Ala} + 0.044A_{U4} - 0.009A_{Cho} + 0.086A_{Crn} + 0.091A_{Tau} - 0.003A_{Gly} + 0.023A_{GAA} - 0.050A_{Cm} - 0.016A_{Lac} - 0.066A_{Thr} - 0.077A_{FuroylGly} + 0.010A_{Tyr} - 0.011A_{His}$$

$$Y_{pr}[2^{nd}T] = 0.282 - 0.037A_{U1} - 0.026A_{U2} - 0.029A_{Leu} - 0.063A_{Ile} - 0.038A_{4-DEA} - 0.054A_{U3} - 0.032A_{3-HBA} + 0.002A_{4-DTA} + 0.031A_{Ala} + 0.021A_{U4} + 0.017A_{Cho} - 0.006A_{Crn} - 0.045A_{Tau} + 0.036A_{Gly} + 0.027A_{GAA} - 0.009A_{Cm} - 0.002A_{Lac} + 0.005A_{Thr} - 0.033A_{FuroylGly} + 0.048A_{Tyr} + 0.060A_{His}$$

$$Y_{pr}[3^{rd}T] = 0.268 + 0.060A_{U1} + 0.060A_{U2} + 0.023A_{Leu} + 0.059A_{Ile} + 0.045A_{4-DEA} - 0.030A_{U3} + 0.038A_{3-HBA} + 0.047A_{4-DTA} + 0.036A_{Ala} - 0.017A_{U4} + 0.024A_{Cho} - 0.049A_{Crn} - 0.030A_{Tau} + 0.017A_{Gly} + 0.006A_{GAA} - 0.010A_{Cm} + 0.019A_{Lac} + 0.052A_{Thr} + 0.021A_{FuroylGly} - 0.015A_{Tyr} - 0.003A_{His}$$

Figure 3.8: Sets of PLS-DA b-coefficients for each class: a) NP, b) 1st T, c) 2nd T and d) 3rd T and corresponding 21-variable discriminant equations (below). A_a : area of peak representative of metabolite a ; variables ordered by descending values of b-coefficients. $y_{pr}[class]$: predicted value reflecting sample class; $y_{pr}=1$ or $y_{pr}=0$ respectively if sample predicted to belong to the class indicated or not. GAA: guanidoacetate; FuroylGly: furoylglycine; Crn: creatinine, 4-DTA: 4-deoxythreonic acid, 4-DEA: 4-deoxyerythronic acid, Lac: lactate, Ui: unassigned i, 3-letter codes used for amino acids.

lipids oxidation and ketone bodies production and U4 (putatively acetyl-carnitine) in the 1st T, a relationship between U1 and U2 (possibly bile acids) and BCAA leucine and isoleucine in 2nd T, and finally a metabolic relationship between U1, U2, 4-DEA and 4-DTA with ketone bodies production in the 3rd T.

3.3.2 Proposed metabolic interpretation of the urinary changes found across pregnancy

The significant changes observed in maternal urine composition throughout pregnancy relate mainly to amino acids and some derivatives, choline, carnitine and 3-HBA, creatinine and four important unassigned resonances. A schematic representation of the metabolic pathways adaptations with and throughout pregnancy is shown in Figure 3.9. Most observations fit broadly with the existing knowledge of the biochemical phenomena accompanying pregnancy (King, 2000, Herrera, 2002, Hod *et al.*, 2008, Creasy *et al.*, 2009). In addition, changes newly found for choline, creatinine, 4-DEA and 4-DTA, furoylglycine, GAA, 3-HBA and lactate, were registered here, to our knowledge for the first time in connection to pregnancy.

Amino acid and their derivatives and lipids metabolism

Pregnancy is usually accompanied by selective aminoaciduria, the exact mechanism of which is not fully known (Hyttén, 1973, Creasy *et al.*, 2009). Moreover, pregnancy is known to be accompanied by increased placental transfer of amino acids, favoring nitrogen conservation for fetal growth (King, 2000, Hod *et al.*, 2008), and thus leading to diminished circulating levels of many amino acids in plasma, a condition known as hypoaminoacidemia. This is also related to an underuse of many amino acids for maternal gluconeogenesis, particularly later in gestation, making way for alternative substrates such as glycerol (Hod *et al.*, 2008).

Lower plasma levels of alanine and other urea cycle amino acids (citrulline, ornithine and arginine) have been reported throughout pregnancy associated to a slowing down of urea synthesis (King, 2000). In spite of the expected redirection of alanine and other amino acids to the fetus, their urinary excretion is typically increased, consistently

with the findings of this work, an aspect which has been proposed to reflect impaired renal reabsorption rather than a saturation phenomenon (Hyttén, 1973, Creasy *et al.*, 2009).

Regarding glycine, its involvement in methionine metabolism might suggest an enhanced production from dimethylglycine, DMG (formed through remethylation of homocysteine to methionine) (Dasarathy *et al.*, 2010). However, this is an unlikely source of glycine in early pregnancy, since the alternative pathway of homocysteine conversion to cysteine through transsulfuration (and then to taurine and/or the reduced form of glutathione, GSH, Figure 3.9) has been proposed as more effective in early pregnancy, in relation to the non-pregnant state (Dasarathy *et al.*, 2010). This is indeed consistent with the observed early increase in excreted choline, possibly due to its underuse in the remethylation pathway to produce DMG, as well as with the trend for higher taurine levels in the 1st T (although with p -value>0.05 the increasing trend is visible in Figure 3.6). It has also been suggested that the preferential activation of homocysteine remethylation occurs in later gestation, which is connected with the choline/DMG pathway or demethylation of tetrahydrofolate (THF) (Dasarathy *et al.*, 2010). The increasing trend seen for choline across the whole pregnancy period (shown in Figure 3.6), although not statistically significant, suggests that the THF pathway may be favored, possibly in connection to folic acid diet supplementation (confirmed for 76% of the pregnant women, information not available for the remaining). Furthermore, the concomitant lower taurine levels, in 2nd and 3rd T, are in agreement with the above mentioned redirection of the methionine pathway.

Regarding branched amino acids (BCAA) leucine and isoleucine, their variations differ from those reported previously (Hyttén, 1973, Creasy *et al.*, 2009), both increasing significantly in the 3rd T. BCCA oxidation is known to decrease in late pregnancy thus increasing BCAA availability for the fetus (King, 2000, Hod *et al.*, 2008) and, similarly to other amino acids, this seems to lead to their enhanced excretion. Through integrals correlations, or STOCSY, it was found that leucine and isoleucine integrals were consistently inter-correlated in all groups (r 0.68, 0.68, 0.63 and 0.85 in NP, 1st, 2nd and 3rd T, respectively), reflecting their common pathway involvement. In addition to this, in the 1st T, isoleucine showed correlations to carnitine (r 0.72) and U4 (r 0.73), the latter two resonances also found to correlate to 3-HBA (r 0.74 and 0.93, respectively for carnitine and U4). This hinders a particular importance of isoleucine in lipid oxidation and ketone

body synthesis, early on in pregnancy even though increased maternal fat deposition and lipogenesis are known to take place preferentially, at that stage (Herrera, 2002).

Carnitine plays an essential role in lipid metabolism as it enables the entrance of long-chain fatty acids into the mitochondria where they undergo β -oxidation producing ketone bodies. Enhanced lipid oxidation is known to occur later in gestation (Herrera, 2002). This enhancement is consistent with the marked decrease of excreted carnitine here found in the 3rd T, suggesting its increased use and consistently with previous reports (Cho and Cha, 2005), and with the accumulation of urinary 3-HBA. In the same trimester, the latter compound was found to correlate with 4-DEA and 4-DTA (r 0.73 and r 0.64, respectively), suggesting a possible relationship of these threonine catabolic products (Appiah-Amponsah *et al.*, 2009, Ellis *et al.*, 2012) to ketone body production, consistently with the ketogenic nature of the amino acid.

Regarding amino acid derivatives, changes were noted for the first time to our knowledge, in the excretion of glycine derivatives guanidoacetate, or GAA, and furoylglycine. Higher excretion of GAA have been previously associated to the accumulation of arginine in the urea cycle (Brosnan and Brosnan, 2010), thus possibly relating to the expected slowing down of that pathway in pregnancy (Figure 3.9). This possible involvement in the urea cycle is consistent with a correlation found for GAA with alanine (r 0.61) in the 1st T. In relation to furoylglycine, a conjugation product of glycine with 2-furoic acid, it is well known that this compound relates strongly to diet, having been related to the consumption of chocolate and heated juices (Kumps *et al.*, 2002). Its relation to diet was confirmed in this work, by its significant increase in the non-fasting test group (shown in section 3.2, Table 3.2) so that it is possible, therefore, that the increase (0.96 effect size) in furoylglycine in the 3rd T may arise from a spurious diet effect. However, the significant decrease in the 1st T (*ca.* 2-fold, $p \approx 10^{-5}$, Table 3.3) allows for the suggestion that a specific relationship of furoylglycine to pregnancy may exist, possibly in relation to lipid metabolism as is the case of other acyl-glycines (Ombrone *et al.*, 2011, Dunn *et al.*, 2012). Finally, still in relation to energy metabolism, it is interesting to note that the documented tendency for increased excretion of glucose in pregnancy (Hyttén, 1973, Creasy *et al.*, 2009) was not detected as significant in the specific conditions of this study, probably indicating its reduced magnitude compared to the impact of other effects such as inter-subject variability and non-fasting. Furthermore, other changes found previously in

lactose (Hyttén, 1973) were not observed in this study, whereas lower abundance sugars (ribose, fucose, fructose) and vitamins (nicotinic acid, folic acid), also expected to change, could not be detected.

Creatinine metabolism

Creatinine excretion was found to decrease significantly with pregnancy, for the 1st time to our knowledge, in spite of earlier evidence that 24 hour urinary creatinine excretion was not altered during pregnancy (Gallery *et al.*, 1996). Creatinine is formed from creatine, which is obtained either from diet or from the conversion of GAA (Figure 3.9), its levels in serum known to decrease throughout pregnancy due to increased glomerular filtration rate (Creasy *et al.*, 2009). The observed decrease in creatinine excretion may reflect the latter effect, however, other variables may contribute to it, namely a possible alteration of creatine synthesis in relation to urea cycle defects (Brosnan and Brosnan, 2010), or the effect of diet (Solanky *et al.*, 2005, Stella *et al.*, 2006). In relation to the latter, a relative decrease in excreted creatinine was indeed observed in samples collected non-fasting (Table 3.2). However, since all samples in this main study were collected in non-fasting conditions, the consistent decrease across pregnancy may also reflect gestation itself.

Important unassigned spin systems

Within the four unassigned resonances seen to change throughout pregnancy, U1 (0.55, s and 0.78 ppm, m) and U2 (0.63 ppm, s) stand out as highly significant. The possibility of these peaks arising from bile acids was investigated through comparison with the spectra of standard compounds (namely, cholic acid, glycocholic acid, chenodeoxycholic acid, deoxycholic acid and glycochenodeoxycholic acid), however, no conclusive assignments could be achieved. Both resonances are positively correlated to each other, in all subject groups but, since their relative intensities are variable, it may be concluded that they originate from metabolically related entities rather than from one single compound. In the 3rd T, U2 further correlates to 3-HBA (r 0.74), which suggests a possible relation of the unassigned singlet to ketone body formation. Regarding U3 (1.18 ppm, 1.45 and 3.82 ppm), correlations were observed in 3rd T with leucine (r 0.64) and isoleucine (r 0.85), suggesting that the compound may be involved in the expected alterations of BCAA oxidation later in gestation. In relation to U4 (putatively acetyl-

carnitine), the noted positive correlations with 3-HBA and isoleucine (r 0.93 and r 0.73, respectively) also suggest a possible involvement in ketone body synthesis.

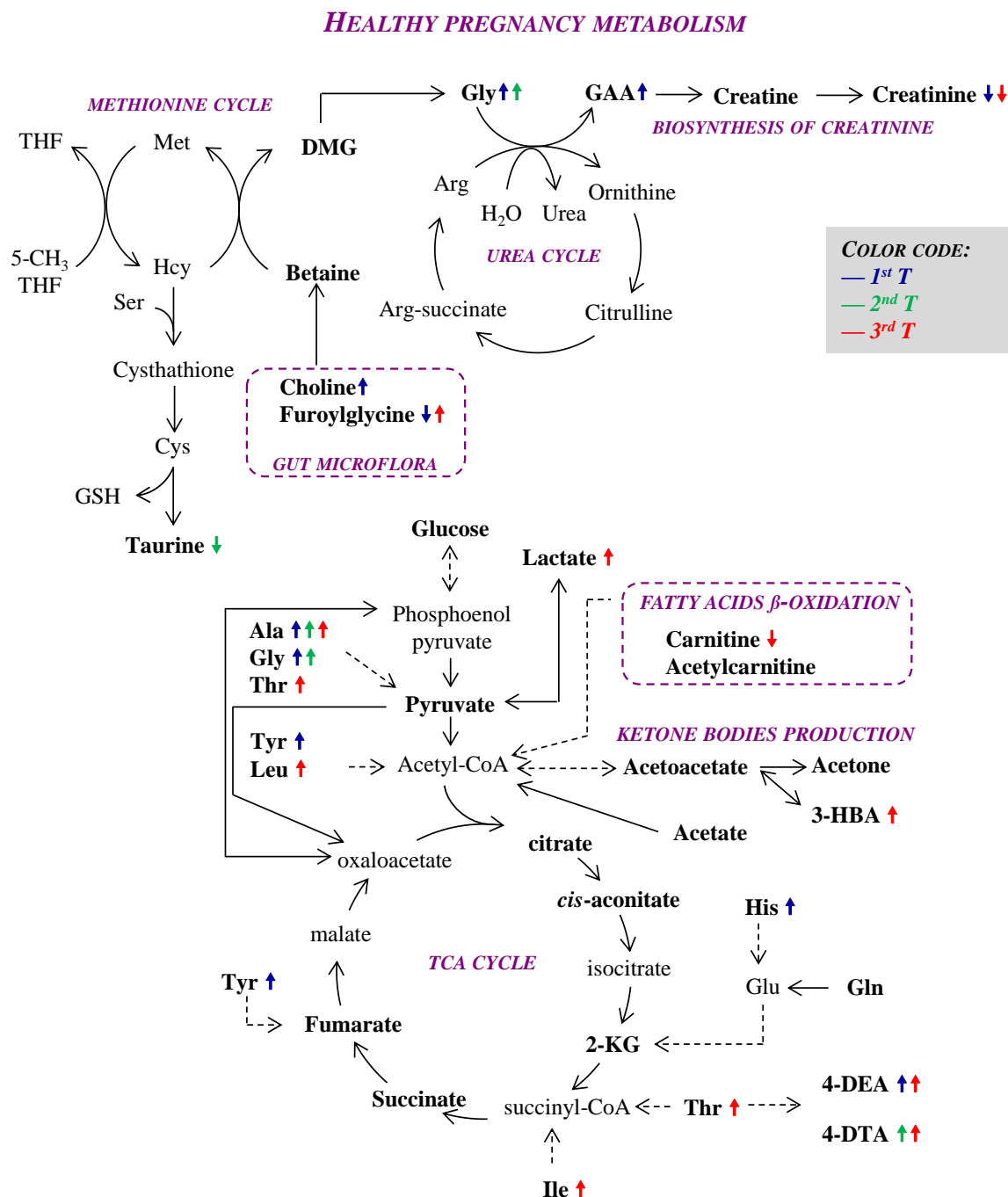


Figure 3.9: Schematic representation of the metabolic pathways adaptations with and throughout healthy pregnancy. Metabolites in bold were identified in ¹H NMR spectra of maternal urine. Amino acids in 3-letter code; 4-DEA: 4-deoxyerythronic acid, 4-DTA: 4-deoxythreonic acid, 3-HBA: 3-hydroxybutyrate, DMA: dimethylamine, DMG: dimethylglycine, GAA: guanidoacetate, Hcy: homocysteine, GSH: glutathione, THF: tetrahydrofolate.

3.3.3 Summary of changes across pregnancy

To sum up, the general increase in amino acid excretion confirmed the expected aminoaciduria accompanying pregnancy, with leucine and isoleucine being more significantly excreted than expected based on the literature, possibly in connection to a marked slowing down of BCAA oxidation in the 3rd T. STOCY analysis suggested that isoleucine was also somehow involved in lipid oxidation (through correlation with carnitine) and ketone body synthesis (through correlation with 3-HBA), in the 1st T. Threonine and its degradation products, 4-DEA and 4-DTA were found to correlate with 3-HBA, possibly in the connection to ketogenesis. The newly noted change in choline was suggested to relate, together with taurine, to methionine metabolism and preferential homocysteine methylation through THF, later in pregnancy. Furthermore, the over excretion of glycine derivative GAA was suggested to relate to the urea cycle deregulation known to accompany pregnancy, whereas another glycine derivative often associated with diet, furoylglycine, was proposed to have a possible particular relationship to pregnancy, due to its marked change in pregnant women, compared to the NP state. A similar proposal was advanced in relation to creatinine, in spite of its concomitant clear dependence on diet (creatinine was found to decrease in samples collected non-fasting). Finally, it is important to note the marked dependence of unassigned resonances U1 (singlet at 0.55 ppm) and U2 (singlet at 0.63 ppm), throughout pregnancy for which a significant gestational age dependence was found, albeit the presently hampered knowledge as to their exact origin.

4. Second trimester maternal urine for the diagnosis of fetal disorders and prediction of poor pregnancy outcomes

In this chapter, the metabolic impact of FM, CD and later developing PTD, PE, IUGR and PROM on the composition of maternal urine was investigated. A novel variable selection method was employed to more efficiently extract information from urine NMR spectra. In addition, a preliminary external validation for selected cases was performed and, in the case of the *pre*-PTD, comparison of NMR and UPLC-MS data was carried out.

4.1 Impact of diagnosed fetal disorders on maternal urine composition

4.1.1 Fetal malformation (FM) cases

For the study of FM, maternal urine collected for women carrying malformed fetuses, diagnosed by ultrasound and undergoing amniocentesis, was compared with that of controls (healthy women with healthy pregnancies). Sample numbers and metadata are presented in Section 2.1.2 b), Tables 2.2 and 2.3 in Experimental Section.

Following a preliminary study, using a smaller cohort and different data treatment (bucketing and no variable selection) published in an initial stage of this work (Diaz and Pinto *et al.*, 2011), some changes were expected in maternal urine of women carrying malformed fetuses. The work described below implies the use of a larger cohort and improved analytical strategies.

The visual comparison of the average ^1H NMR spectra of urine of healthy (controls) and FM cases revealed some subtle changes mainly in the aliphatic region, namely in 3-hydroxybutyrate (3-HBA), alanine, acetone, acetoacetate, succinate, and aromatic region, namely in 4-hydroxyphenylacetate (4-HPA) and tyrosine (Figure 4.1).

PLS-DA was applied to the NMR datasets to aid more detailed comparison of control and FM cases, although the obtained model was found to have no predictive power, as given by the low Q^2 value (0.251) and MCCV results showing high overlap of Q^2 distributions for true and permuted models and poor ROC, low classification rate (69%), specificity (75%) and sensitivity (49%) rates (Figure 4.2, Table A-III.2, Annex III). The lack of model robustness probably arises from the complex nature of urine, which typically contains information reflecting several different factors, such as diet, drugs, environment (Bouatra *et*

al., 2013), these factors having comparable (or possibly even higher) impact in urine composition to the changes found for FM. Thus, it was necessary to devise a way to retrieve the spectral data associated with sample class, in this case control *vs.* FM. This was carried out by applying a variable selection methodology to remove irrelevant, unreliable and noisy variables. The development and validation of a variable selection method is described in detail throughout the next section.

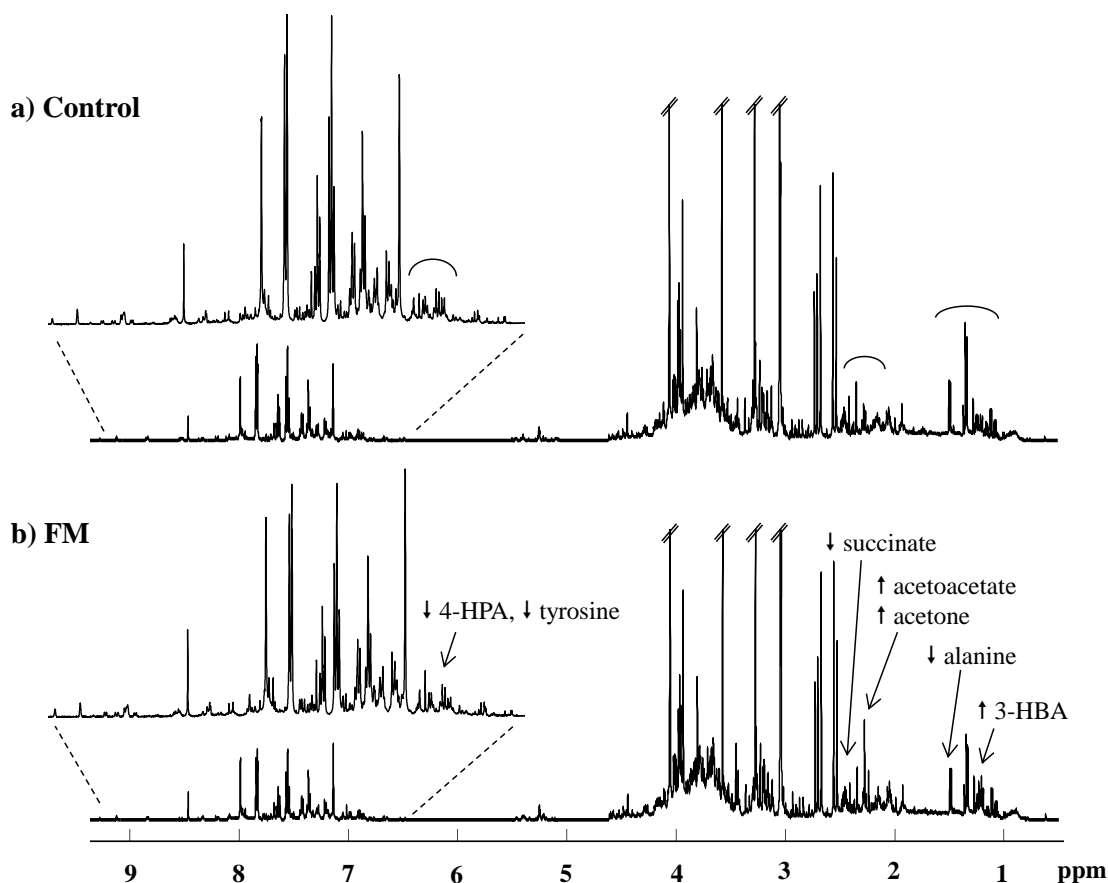


Figure 4.1: Average 500 MHz ^1H NMR spectra of 2nd T maternal urine of a) controls ($n=84$) and b) women carrying malformed fetuses ($n=35$). Curved lines indicate regions where changes are noted and arrows point at specific changing metabolites.

4.1.1.1 Development and validation of a variable selection methodology

To develop the variable selection method for NMR data, two PLS-DA parameters were considered due to their complementarity, namely the VIP and b-coefficient values. b-coefficients values derive from the discrimination function and reflect the discriminant and predictive capability of each variable i.e. their contribution for samples classification. In

addition, VIP values reflect the importance of each variable to the projection in the low dimensional space i.e. the relevance of each variable for samples geometric separation. Both parameters are based on the multivariate method PLS-DA, thus both taking into account the concomitant changes and interactions between variables/metabolites. Because these interactions are determinant for the definition of the overall profile, univariate methods were not used for the development of this variable selection methodology, as these would only take into account individual metabolites changes.

In this work, four variable methodologies were tested. The first method explored (method i) was based on the selection of variables with $VIP > 1$, these variables being by definition relevant for class separation. The method had been previously reported for the study of cancer through blood plasma metabolomics (in an animal model) (Sun *et al.*, 2012) and foodstuff adulteration (specifically the contamination of commercial spices with Sudan dyes colorants) (Di Anibal *et al.*, 2011). The second and third methods explored (respectively named as methods ii) and iii)) were based on signal uncertainty using a jack-knife approach obtained by cross-validation (Andersen and Bro, 2010). For these methods, the VIP- and b-coefficient standard errors were obtained, as determined by cross-validation, respectively VIP_{cvSE} and b_{cvSE} . Method ii) consisted in selecting variables with $VIP/VIP_{cvSE} > 1$ whilst the method iii) was based on selecting variables with $|b/b_{cvSE}| > 1$. Lastly, the interception of all previous variable selection methods was attempted and named as method iv), corresponding to the selection of variables with simultaneously high VIP and low VIP- and b-coefficient standard errors. The principle of data interception has been used before in only one instance (study of nonalcoholic fatty liver disease (Quintás *et al.*, 2012)) although based on different criteria.

Upon application of each variable selection method, PLS-DA were recalculated using the resulting data subsets and resubmitted to MCCV, with the resulting model quality parameters, validation results and number of selected variables shown in Table A-III.2, Annex III. The best models were obtained with method iv), the interception method, which yielded good model parameters (Q^2 , Q^2_{median} , classification rate, sensitivity and specificity). Figure 4.2 illustrates the performance of this variable selection method when applied to the study of FM. The significant reduction in samples dispersion can be seen in the PLS-DA score plots after variable selection (Figure 4.2a,d) as well as the noteworthy improvement in Q^2 distribution, with significantly reduced overlap of true and permuted models (Figure

4.2b,e) and improvement in specificity and sensitivity, visible through ROC plots (Figure 4.2c,f). Moreover, a significant improvement in classification rate (CR), sensitivity and specificity (respectively 84%, 74% and 87%) was achieved, compared to those obtained using the original dataset (respectively 69%, 49% and 75%).

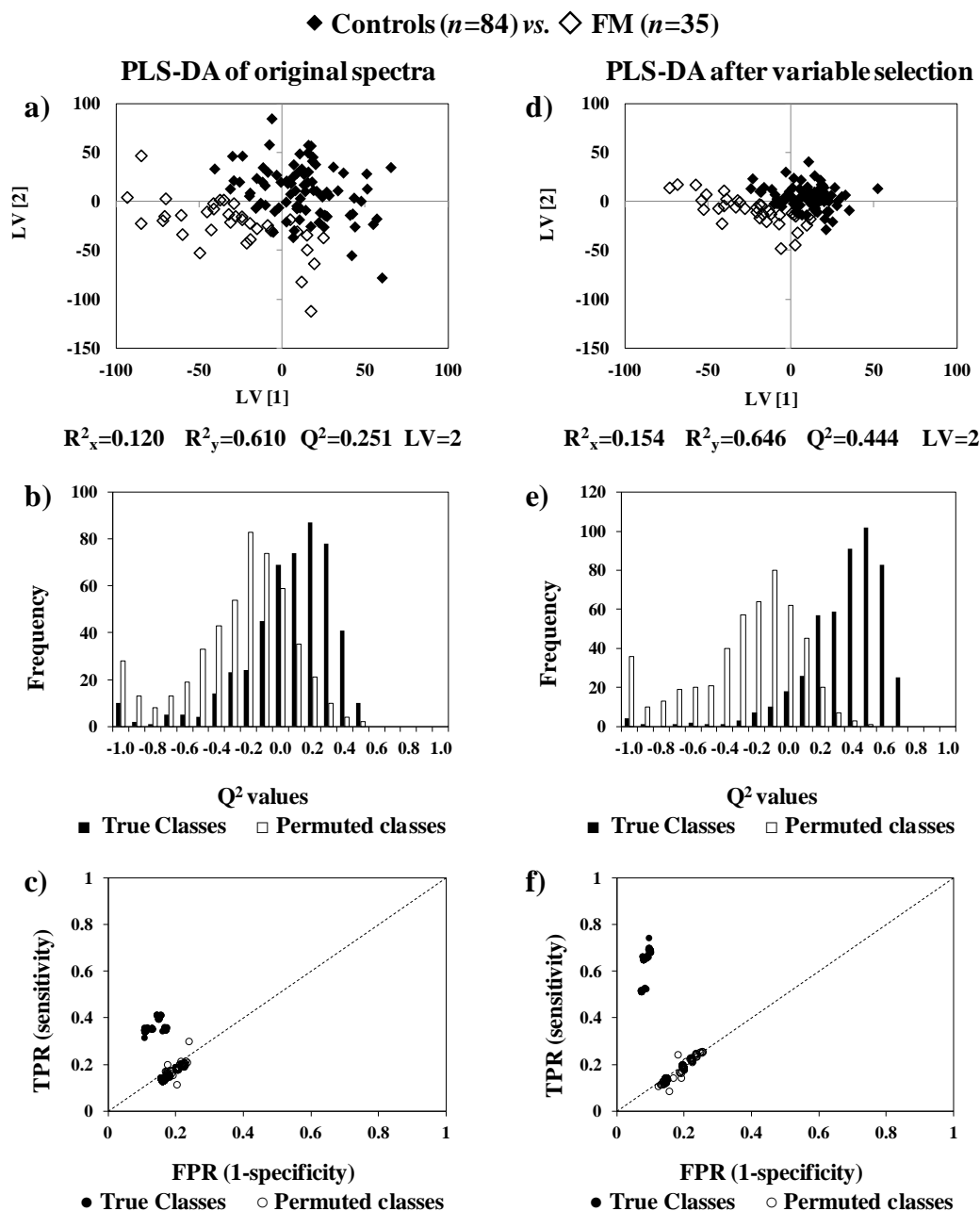


Figure 4.2: PLS-DA scores plots (a, d), Q^2 distributions (b, e), and ROC plots (c, f) of true and permuted models obtained for the original ^1H NMR spectra of urine of controls (◆, $n=84$) and FM (◇, $n=35$) and for the same spectra after variable selection using intersection method iv). TPR, true positive rate; FPR, false positive rate.

This selection method enabled the most drastic variable reduction of the original dataset without compromising the model's predictive power, as the best result was consistently obtained using this variable selection method. The significant data reduction poses two major advantages, the first being that only spectral regions that accurately define the condition (thus, the potential metabolic profile) are retrieved. The second advantage consists of the lowering of the variables/samples ratio, thus reducing the risk of model overfitting. All NMR results presented from this point onwards were obtained using this variable selection methodology.

To retrieve the urinary profile on the basis of the PLS-DA separation, the average NMR spectrum of the control and FM groups were compared along with indication of the subset of spectral variables selected. An example of this spectra comparison is shown in Figure 4.3, for the 0.5-4.6 ppm spectral region. In this figure, the gray dots under the bottom spectra indicate variables selected, and the insets note that relevant variables may

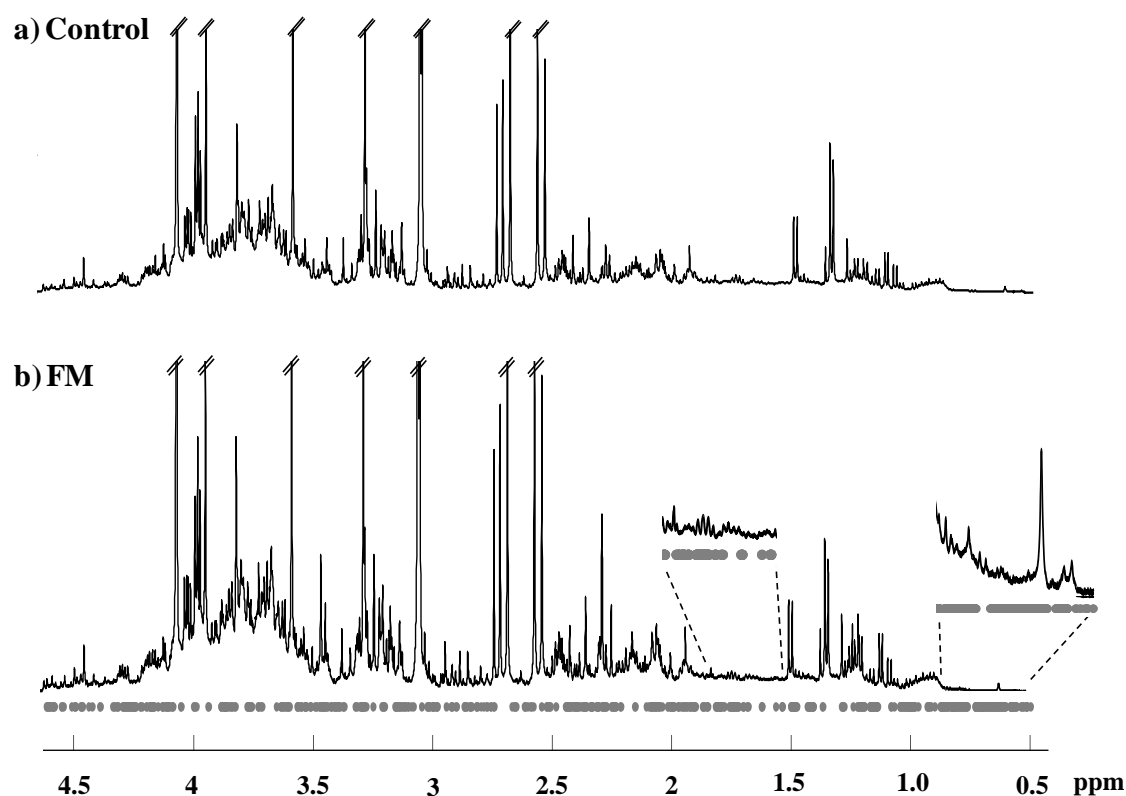


Figure 4.3: Average ¹H NMR spectra of a) control and b) FM groups (0.5-4.6 ppm) with indication of spectral variables, i.e. data points/gray dots under bottom spectra, selected by variable selection method iv.

fall in regions with low signal intensity but confirmed to always bear visible resonances at the appropriate level of vertical expansion. This demonstrates the importance of low intensity, and often unassignable, NMR signals in incorporating the potential fingerprint of a given disorder.

4.1.1.2 Unveiling urinary metabolic signatures of general FM and central nervous system (CNS) malformations

The group of selected variables supporting the separation seen in the PLS-DA score plot in Figure 4.2d enabled a NMR signature to be identified for FM cases. This signature comprises changes in 17 assigned compounds, 7 unassigned spin systems (U_i) (singlet at 0.63 ppm, doublet at 1.08 ppm, doublet at 1.15 ppm, singlets at 1.99 ppm, 2.05 ppm, 2.06 ppm and 2.39 ppm) and several unassigned spectral regions (defined as spectral regions comprising overlapped unidentifiable resonances). Each metabolite, unassigned compound or spectral region were tested for statistical (p -value) and biological significance (effect size) (Table 4.1, left). In this table, metabolites that may originate from diet (either directly e.g. xylose, 3-methylhistidine, or through their interaction with the gut microflora e.g. 4-hydroxyphenylacetate (4-HPA), *scyllo*-inositol) are marked with ^b.

Overall, 18 signals were found to vary with statistical relevance ($p < 0.05$). PLS-DA was repeated considering only this set of significant integrals ($p < 0.05$), the resulting scores plot and MCCV results being shown in Figure 4.4. Through this figure, it becomes clear that no reliable group separation was achieved (with high overlap of samples in the scores plot and of Q^2 values distribution of true and permuted models) when only significant integrals were taken into account. This finding supports the idea that using full spectral resolution of NMR spectra (including unassigned signals and signals with no individual significance, $p > 0.05$) is of paramount importance for an effective disease signature to be achieved.

Table 4.1: Metabolite/resonance changes in maternal urine of FM vs. controls and CNS vs. other FMs.

^aChemical shifts selected by variables selection; TOCSY or STOCSY correlated peaks in square brackets; s: singlet, d: doublet, t: triplet, q: quartet, dd: doublet of doublets, m: multiplet, br: broad. ^bMetabolites related to diet and/or gut microflora; U10 correlates to trigonelline, suggesting a dietary dependence, as previously found in non-fasting (Chapter 3). ^cChanges specific of CNS malformations. ^dConsistent with previous findings published as result of this work (Diaz and Pinto *et al.*, 2011). ^ePeaks with minimal variations. Ui: unassigned compound i by appearance order. 4-DTA: 4-deoxythreonic acid, DMG: dimethylglycine, 3-HBA: 3-hydroxybutyrate, 2-HIBA: 2-hydroxyisobutyrate, 4-HPA: 4-hydroxyphenylacetate, IS: indoxyl sulphate, 2-KG: 2-ketoglutarate, NMND: *N*-methylnicotinamide, TMA: trimethylamine. Only *p*-values<0.05 are shown.

FM <i>n</i> =35 vs. controls <i>n</i> =84			CNS <i>n</i> =12 vs. other FMs <i>n</i> =23		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)	Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
3-HBA	1.2 d, 4.15 m	$\uparrow(0.80, 3.16 \times 10^{-2})$	2-HIBA ^{c, e}	1.36 s	$\uparrow(0.05)$
3-Me-histidine ^b	3.75 s, 7.15 s	$\uparrow(0.50, 6.24 \times 10^{-3})$	2-KG ^c	2.45 t, 3.01 t	$\uparrow(0.38)$
4-HPA ^b	6.86 d	$\downarrow(-0.47, 3.2 \times 10^{-2})$	3-Me-histidine ^b	3.75 s, 7.15 s	$\uparrow(0.21)$
Acetoacetate	2.28 s, 3.45 s	$\uparrow(1.00)$	4-DTA ^c	1.23 d, 4.12 m	$\downarrow(-0.26)$
Acetone	2.23 s	$\uparrow(0.30)$	4-OH-hippurate ^b	6.98 d, 7.76 d	$\uparrow(0.47)$
Acetyl-carnitine	2.15 s, 3.19 s	$\uparrow(0.42)$	Acetyl-carnitine	2.15 s, 3.19 s	$\uparrow(0.59, 2.1 \times 10^{-3})$
Alanine	1.48 d	$\downarrow(0.11)$	Alanine ^c	1.48 d	$\uparrow(0.44)$
<i>cis</i> -aconitate ^d	3.12 d	$\uparrow(0.20)$	Allantoin ^c	5.39 s	$\uparrow(0.52)$
Isoleucine ^d	1.01 d	$\uparrow(0.76, 9.94 \times 10^{-4})$	Acetate ^c	1.93 s	$\downarrow(-0.91, 3.7 \times 10^{-3})$
Lactose	4.46 dd	$\uparrow(0.33)$	Carnitine ^c	3.23 s, 2.44 dd	$\uparrow(0.92, 2.3 \times 10^{-2})$
Lysine	1.73 m, 1.91	$\downarrow(-0.35, 1.4 \times 10^{-2})$	Choline ^{b, e}	3.20 s	$\downarrow(-0.02)$
NMND ^d	4.48 s, 8.90 d,	$\uparrow(0.67, 7.1 \times 10^{-4})$	Citrate ^c	2.54 d, 2.67 d	$\uparrow(0.73)$
	8.97 d, 9.28 s		DMG ^c	2.93 s	$\uparrow(0.72)$
<i>Scyllo</i> -inositol ^b	3.36 s	$\downarrow(-0.40, 3.92 \times 10^{-2})$	Furoylglycine ^b	6.65 q, 7.70 d	$\downarrow(-0.65)$
Succinate	2.41 s	$\downarrow(-0.30)$	Glutamine ^c	2.15 m, 2.47 m	$\downarrow(-0.64, 3.4 \times 10^{-2})$
Threonine ^d	4.26 dd	$\uparrow(0.56, 1.72 \times 10^{-2})$	Isoleucine	1.01 d	$\uparrow(0.73)$
Tyrosine	6.91 d	$\downarrow(-0.38, 4.04 \times 10^{-2})$	IS ^b	7.70 d	$\uparrow(0.45)$
Xylose ^b	4.58 d, 5.21 d	$\uparrow(0.35)$	Lactose	4.46 dd, 5.25 d	$\uparrow(0.95)$
U1 ^d	0.63 s	$\uparrow(0.51, 3.87 \times 10^{-4})$	Leucine ^c	0.97 t	$\uparrow(0.82)$
U2	1.08 d [2.50,	$\downarrow(0.76)$	<i>Scyllo</i> -inositol ^b	3.36 s	$\uparrow(0.40)$
	3.55, 3.72]		Succinate ^c	2.41 s	$\uparrow(0.48)$
U3	1.15 d [3.45,	$\downarrow(-0.40, 1.83 \times 10^{-2})$	Trigonelline ^b	4.44s, 8.84br,	$\downarrow(-0.54)$
	3.55]			9.12 s	
U4	1.99 s	$\downarrow(-0.40)$	TMA ^b	2.89 s	$\uparrow(0.29)$
U5	2.05 s	$\uparrow(0.74, 2.49 \times 10^{-3})$	Valine ^c	0.99 d, 1.04 d	$\uparrow(0.39)$
U6 ^d	2.06 s	$\uparrow(0.78, 1.47 \times 10^{-3})$	U2	1.08 d [2.50,	$\uparrow(0.88)$
U7	2.39 s	$\uparrow(0.82, 4.84 \times 10^{-2})$		3.55, 3.72]	
Unassigned spectral regions			U3	1.15 d [3.45,	$\downarrow(-0.39)$
0.65-0.73		$\uparrow(0.84)$		3.55]	
0.77-0.83		$\uparrow(0.62)$	U8 ^d	2.87 s	$\uparrow(0.74)$
1.40-1.43		$\uparrow(0.14)$	U9 ^c	2.91 s	$\uparrow(1.04, 3.5 \times 10^{-3})$
3.41-3.42		$\downarrow(-0.17)$	U10 ^b	4.40 s, 8.78 d	$\downarrow(-0.58)$
3.85-3.87		$\uparrow(0.61, 3.03 \times 10^{-2})$	U11 ^c	6.84 br, 6.86 br,	$\uparrow(0.97)$
5.30-5.32		$\uparrow(0.13)$		6.92 br, 7.31 s	
5.35-5.38		$\uparrow(0.33, 4.09 \times 10^{-2})$	U12 ^c	0.55 s	$\uparrow(0.41)$
6.56-6.57		$\downarrow(-0.37)$	U13 ^c	0.83 s	$\uparrow(0.66)$
8.16-8.22		$\uparrow(0.57, 3.32 \times 10^{-3})$	U14 ^c	0.93 s	$\uparrow(0.19)$
8.34-8.36		$\uparrow(0.52, 5.26 \times 10^{-3})$	U15 ^c	1.10 d	$\downarrow(-0.21)$
			U16 ^{c, e}	1.25 d	$\downarrow(-0.06)$
			U17 ^c	6.81 s	$\uparrow(0.14)$
			U18 ^c	9.36 s	$\downarrow(-0.64)$
			Unassigned spectral regions		
			3.88-3.90		$\downarrow(-0.60, 4.9 \times 10^{-2})$
			+ 15 selected regions, <i>p</i> >0.05, mostly at 0-4.6 ppm		

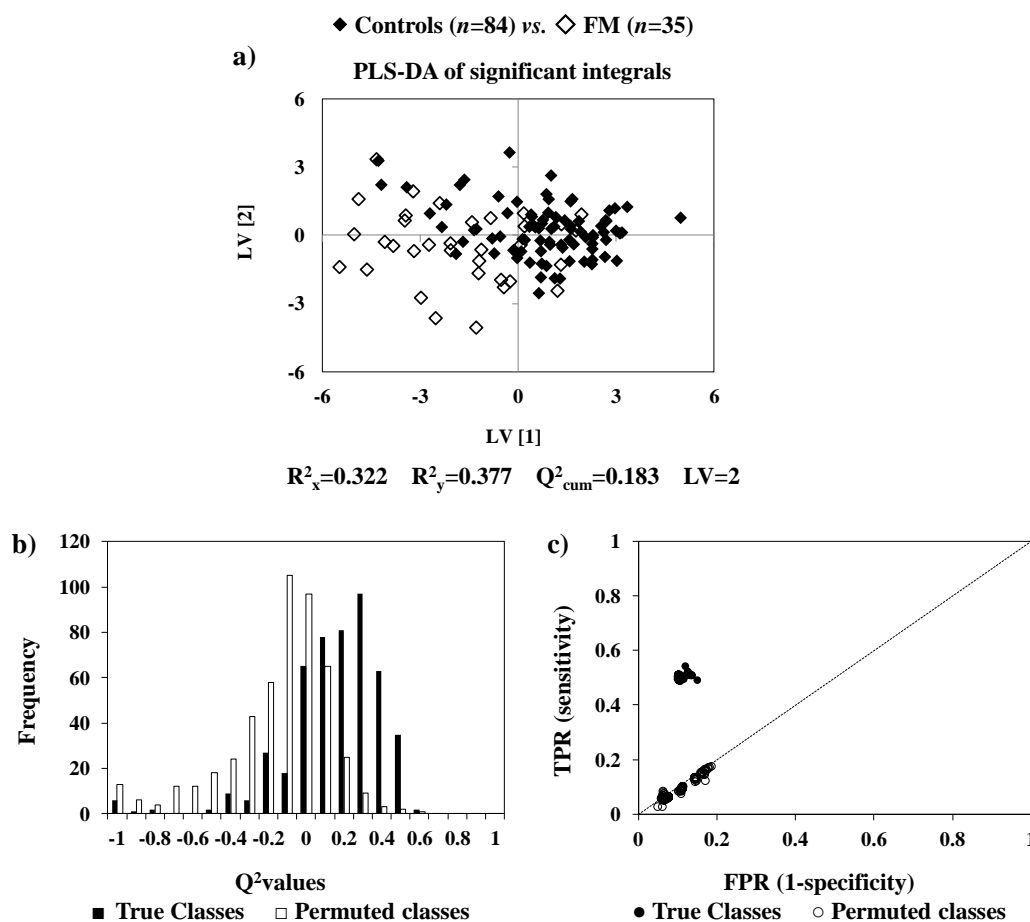


Figure 4.4: a) PLS-DA scores plot, b) Q^2 distributions plot and c) ROC plot of true and permuted classes, given by MCCV of the set of integrals of the statistically relevant ($p<0.05$) metabolites/regions for FM (◇, $n=35$) vs. Controls (◆, $n=84$) (see Table 4.1, left). TPR: true positive rate, FPR: false positive rate.

The possibility of specifying FM type through maternal urine was also considered and, indeed, it was found that CNS malformations were distinguishable from the remaining FM types, as shown in the PLS-DA score plot obtained after variable selection (Figure 4.5a). MCCV of this model revealed high predictive power (Q^2 0.705, and low overlapping of Q^2 distribution of true and permuted models) and robustness (with 87% CR, 83% sensitivity, 90% specificity) as shown in Figure 4.5b,c) and Table A-III.2, Annex III). This result expresses the potential of urine metabolomics to aid and complement clinical determination of FM type.

The metabolic profile of CNS was retrieved, and is recorded in Table 4.1 (right). Changes were also analysed in terms of specificity and some apparently CNS specific

changes were identified. Changes considered specific included reversed variations such as that noted for alanine, this being decreasing in FM vs. controls whilst increasing in CNS vs. other FMs. Specific CNS changes were noted for 13 metabolites and 9 unassigned resonances these being indicated with a ^c in Table 4.1, right.

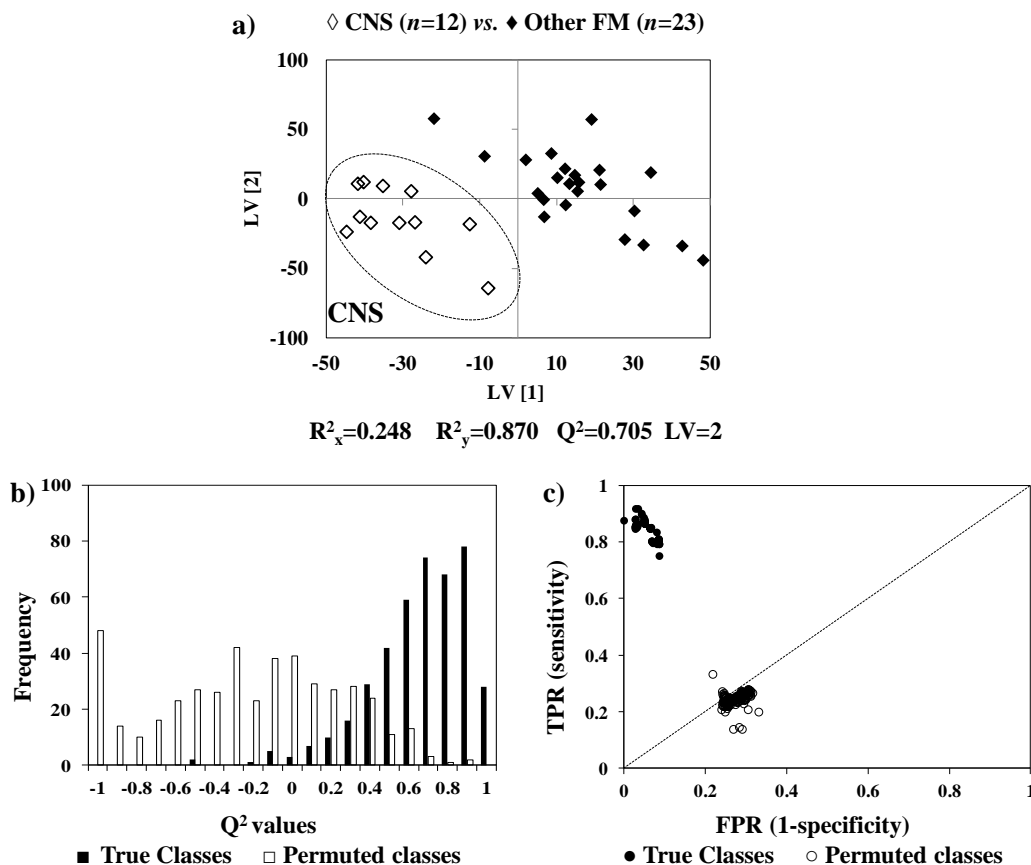


Figure 4.5: a) PLS-DA scores plot, b) Q^2 distributions plot and c) ROC plot of true and permuted classes given by MCCV, obtained for 2nd trimester maternal urine of women carrying fetuses with CNS malformations (◇, $n=12$) vs. other types of FM (◆, $n=23$). TPR: true positive rate; FPR: false positive rate.

To unveil the urinary metabolic signatures of the general FM group and of CNS malformations, a novel circular representation of the selected profile with corresponding importance of each selected variables was devised. In this plot, shown in Figure 4.6, named here as VIP-wheel representation, the inner circle represents an average ¹H NMR spectrum of the control group and each point represents a variable selected with the magnitude of VIP in the radial dimension. Inspection of this plot enables the urinary metabolic signature of FM (shown in red) and that of CNS (shown in orange) to be unveiled, thus illustrating

visually the specific features of the latter disorder. For instance, *N*-methylnicotinamide (NMND) was clearly selected as important for the FM profile with high VIP values in the radial scale but not for CNS. Specific changes for CNS are also visible, namely for citrate, acetate and TMA, which were selected with high VIP values, but not in FM.

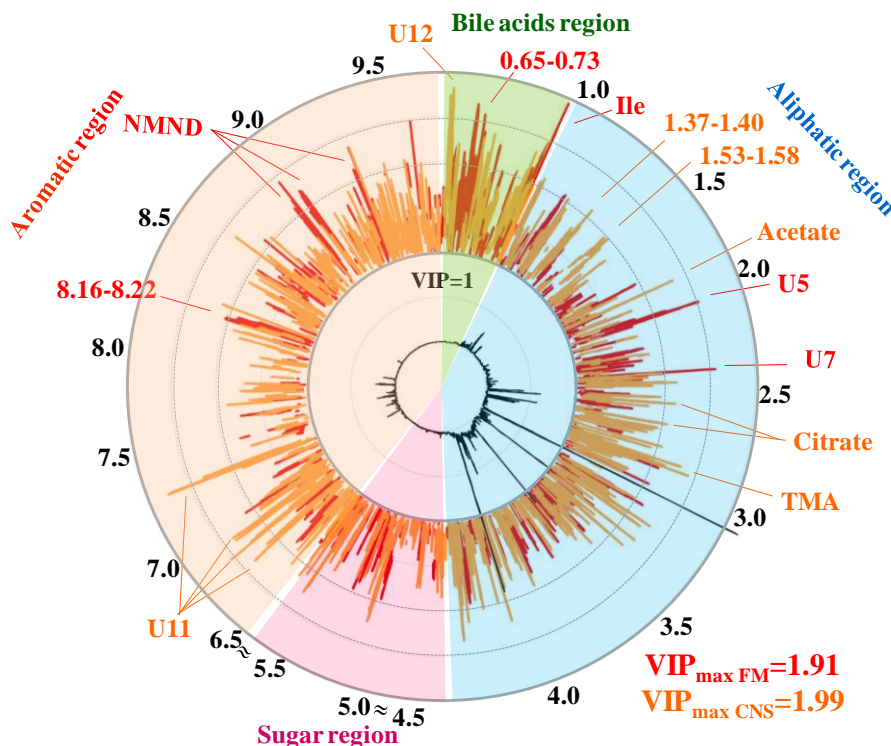


Figure 4.6: VIP-wheel representations of the NMR metabolite signatures obtained for FM (red) overlaid with CNS malformations (orange). The average ^1H NMR spectrum of controls is represented in the inner circle, with the corresponding ppm scale shown in the outer black circle. Each dot represents a variable (spectral data point) selected by the intersection method, positioned in a radial VIP scale according to its VIP value. Some distinguishing features are indicated.

4.1.1.3 Preliminary external validation

The possibility of classifying new samples using the previously developed models is of paramount importance, especially when a clinical application is envisaged. The recommended approach would be to divide samples into training and prediction sets, i.e. samples used to develop and to test the model, respectively. In this work the training phase of the study, consisting of the development of the initial PLS-DA model used for further external validation, was carried out with the initially available samples (84 controls and 35 FM). Subsequently, new samples were collected and 29 new controls and 6 new FM cases

(these including 2 CNS and 4 other types of malformations) became available to test the model. In order to do this, the ^1H NMR spectra of all samples (FM and controls, training and prediction sets) were re-aligned and normalized, as described in the Experimental Section, and MVA was repeated considering only those variables selected in the previous model (this corresponds to the variable subset selected using the intersection method and shown in Figure 4.2 and Figure 4.3) using the same initial FM and control groups (training sets). The scores of the new set of samples (i.e. the prediction set) were predicted and classification was attempted based on the discrimination function and shown in the form of a confusion matrix. The resulting PLS-DA score plots of the training and prediction sets are shown in Figure 4.7 left, along with the corresponding confusion matrix for the new FM and control cases. The same approach was carried out to attempt malformation type classification, using the variable subset determined for CNS vs. other FM types model, shown previously in Figure 4.5. The resulting predicted scores and confusion matrix are shown in Figure 4.7 right.

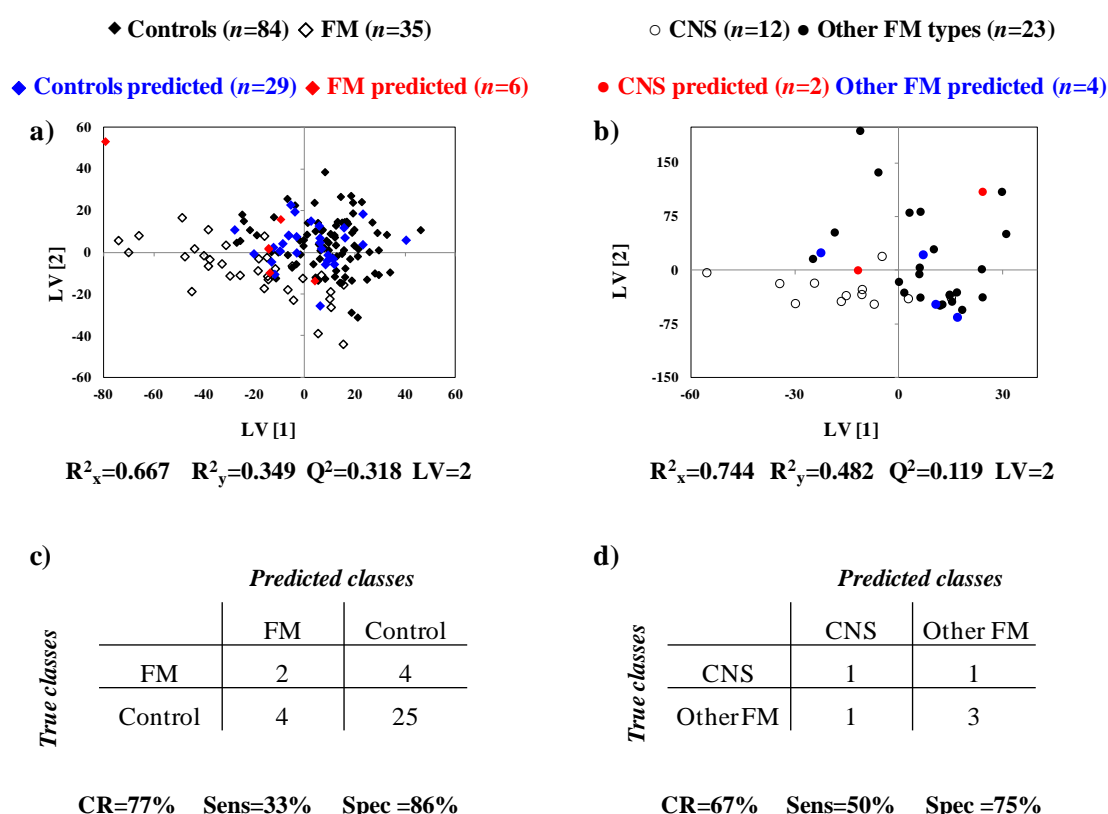


Figure 4.7: Original and predicted score plots and confusion matrices for a,c) FM and controls and b,d) CNS and other FM types. CR: classification rate, Sens: sensitivity, Spec: specificity.

The prediction results obtained the FM *vs.* controls model, show the overlap of the training and prediction sets samples (Figure 4.7a), although the scores corresponding to the predicted FM and control samples (prediction set), respectively red and blue diamonds, are not as separated as samples used in the training set, shown in black opened and dashed diamonds. Moreover, classification results, i.e. classification rate 77%, sensitivity 33% (expressing a very low TPR, i.e. classification of FM cases, or true positives, as only 2 out of 6 were correctly classified) and specificity 86%, computed through the confusion matrix shown in Figure 4.7c. In relation to the CNS *vs.* other FM model, samples from the training and prediction sets are overlapped in the score plot (Figure 4.7c), but no separation trend is visible between CNS and other FM predicted scores, shown in red and blue dots respectively. Moreover, a significant decrease in Q^2 (0.119) was obtained, indicating the low predictive power of the model. Regarding sample classification, poor results were obtained, the classification rate being 67%, sensitivity 50% (as only 1 out of the 2 CNS samples was correctly classified) and specificity 75%.

The results obtained with external validation, show that the model here presented is not yet robust and capable of classifying new samples. We believe that this limitation may be related to two factors. One consists on the unbalanced sample numbers of the groups used to build the models (84 controls, 35 FM; 12 CNS and 23 other FM). This means that the classification of a new sample in the larger groups (in this case controls or other FM) is more probable. A potential solution to this limitations would be to increase the number of samples in the FM and CNS groups, thus using more balanced groups. A second limitation may arise from the initial step of spectral alignment, as it is very dependent on the set of samples being aligned. This means that when performing alignment to different sets of samples it is impossible to assure that each spectral point (variable, 0.0003 ppm width) corresponds always to the same information, i.e. peak intensity, due to peak positional differences caused by spectral alignment. One possible manner to overcome this pitfall would be to use spectral bucketing instead of full resolution data followed by spectral alignment, thus assuring that each bucket window corresponds always to the same spectral information in all samples. Both bucketing and alignment have advantages/disadvantages and these must be weighted for each specific application envisaged. The use of full spectral resolution, after peak alignment provides significant gain in resolution and its use is invaluable to unveil spectral changes often occurring in narrower ranges compared to

bucket widths. On the other hand, the use of spectral bucketing may be advantageous if a clinical application is envisaged since it allows more reproducible and independent (of the sample set) models to be obtained.

4.1.1.4 Proposed metabolic interpretation of urinary changes in general FM and CNS malformations

Regarding the general FM signature, listed in Table 4.1 left, changes were consistent with previously advanced hypotheses regarding the biochemical perturbation occurring in cases of FM. A general scheme of the metabolic pathways affected in FM and CNS is depicted in Figure 4.10, along with those affected in CD and T21 as will be presented and discussed in the following section.

It had been proposed that FM cases show decreased glucose levels in AF (Graça *et al.*, 2010) which reflects the activation of gluconeogenesis and energy production through lipid oxidation and ketone body synthesis (Graça *et al.*, 2010, Diaz and Pinto *et al.*, 2011, Graça *et al.*, 2012). In this work, higher maternal urinary excretion of all three ketone bodies (acetone, acetoacetate and 3-hydroxybutyrate, 3-HBA) was observed. These findings are in agreement with enhanced lipid oxidation, leading to over production and excretion of ketone bodies and production of acetyl-CoA. The resulting acetyl-CoA feeds the TCA cycle in extrahepatic tissues, enhancing its activity, which had been found in FM cases (Graça *et al.*, 2012). This perturbation is further supported by the higher *cis*-aconitate excretion found in this cohort and consistent with a previous exploratory report published as result of this work (Diaz and Pinto *et al.*, 2011). The increase in excreted acetyl-carnitine suggests that an excess of acetyl-CoA is formed, locking carnitine in its acetylated form, also supporting the perturbation in lipid oxidation. Decreased succinate, substrate of the enzyme complex II of the respiratory chain, was found in maternal urine of FM cases. Increased succinate had been previously found in AF (Graça *et al.*, 2010) suggestive of an inadequate use of the respiratory chain for energy production. Thus, the lower excretion of succinate found in this cohort may suggest its possible diversion to the fetus. Relevant changes were also found regarding amino acid excretion, possibly reflecting their selective use as gluconeogenesis substrates, possibly preferentially for alanine, lysine, and tyrosine, which are being less excreted, compared to isoleucine and

threonine, which are being increasingly excreted. Out of these amino acids, alanine, lysine, tyrosine and isoleucine had been found decreased in AF of FM cases due to increased amino acid catabolism to replenish glucose (Graça *et al.*, 2010, Graça *et al.*, 2012). Furthermore, increased isoleucine and threonine had been found in maternal urine of FM in a previous exploratory report resultant of this work, this variation being associated with enhanced gluconeogenesis and TCA cycle demand (Diaz and Pinto *et al.*, 2011). Other significant change found in this work is the higher excretion of NMND, a metabolite involved in nucleotide metabolism and particularly the tryptophan-NAD⁺ pathway (Salek *et al.*, 2007). Increased NMND had been found in smaller cohort (initial work within this thesis, published in (Diaz and Pinto *et al.*, 2011)) being further confirmed in this larger cohort. Furthermore, higher lactose excretion was found in the FM group, this having been seen in healthy pregnant women (Hyttén, 1973, Sachse *et al.*, 2012) but not in relation to prenatal disorders.

CNS malformations seem to be characterized by an abnormal change in the acetyl-carnitine/carnitine balance, reflecting somewhat higher circulating acetyl-CoA levels and an underuse of carnitine, possibly reflecting a relatively lower level of lipid oxidation compared to other FM types. This finding is consistent with previous reports which had shown changes in maternal plasma lipids in CNS cases (Diaz and Pinto *et al.*, 2011). The decrease in acetate suggests its use as substrate for acetyl-CoA synthesis, contributing to enhanced TCA cycle activity. This suggestion is also supported by the increased excretion of intermediates 2-ketoglutarate (2-KG), citrate and succinate. Succinate is further involved in the respiratory chain, which has been found to be impaired in women carrying fetuses with neural tube defects (NTD) (Zheng *et al.*, 2011) and increased levels of succinate have been reported in AF in cases of spina bifida (Groenen *et al.*, 2004). CNS cases also show specific disturbances in branched chain amino acids (BCCA) metabolism (higher excretion of leucine, isoleucine and valine) and of other amino acids (alanine and glutamine) and amino acids derivatives 4-deoxythreonic acid (4-DTA) and dimethylglycine (DMG). Among these changes, the increase of DMG here found may relate to methionine metabolism, known to be altered in NTD cases (Dasarathy *et al.*, 2010) and found to be down regulated in women carrying fetuses with this type of malformations (Zheng *et al.*, 2011). Furthermore, decreased valine and increased

glutamine (the latter variation being in the opposite direction to that found in this work) had been previously found in the maternal serum (Zheng *et al.*, 2011) of women carrying fetuses affected by NTD. Increased excreted allantoin and indoxyl sulphate (IS) suggest oxidative stress in CNS cases, with allantoin being an oxidation product of uric acid and a marker of oxidative stress, and IS having been suggested to mediate the oxidative stress process (Žitňanová *et al.*, 2004, Dou *et al.*, 2007). Moreover, an apparent contribution from gut microflora metabolites was also found in CNS signature, given by changes in 4-hydroxyhippurate, choline (residual change), trigonelline and trimethylamine (TMA) excretion.

4.1.2 Chromosomal disorder (CD) cases

4.1.2.1 Unveiling urinary metabolic signatures of general CD and trisomy 21 (T21)

The metabolic effect of carrying a fetus with a CD was evaluated by comparing ^1H NMR spectra of maternal urine from cases and controls, collected at the time of amniocentesis during the 2nd trimester of pregnancy. Diagnosis was done by determination of the amniocytes karyotype in the cytogenetics laboratory of the FMUC and results were available approximately two weeks after amniocentesis. Sample numbers and metadata are shown in Tables 2.2 and 2.3 (Experimental Section).

An initial PLS-DA of maternal urine of CD and controls, using the original NMR dataset, revealed a separation trend (not shown) between groups, although MCCV resulted in low CR, sensitivity and specificity (respectively 68%, 34% and 73%) along with high overlapping Q^2 values of true and permuted models. As also was the case for FM, this unsatisfactory result may be explained by the low impact of the disease in urine composition, compared with the amount contained in urine. Thus, the variable selection method presented and discussed in Section 4.1.1.1 was applied to CD and PLS-DA was recalculated. This resulted in a clearer distinction of CD and controls cases (Figure 4.8a), although some dispersion was still visible between CD samples possibly due to the heterogeneous nature of the group. Regarding the two outlier samples visible in the score plot, visual inspection of the corresponding spectra revealed high resonances of glucose, whilst one the samples had also high lactate, pyruvate and 2-KG.

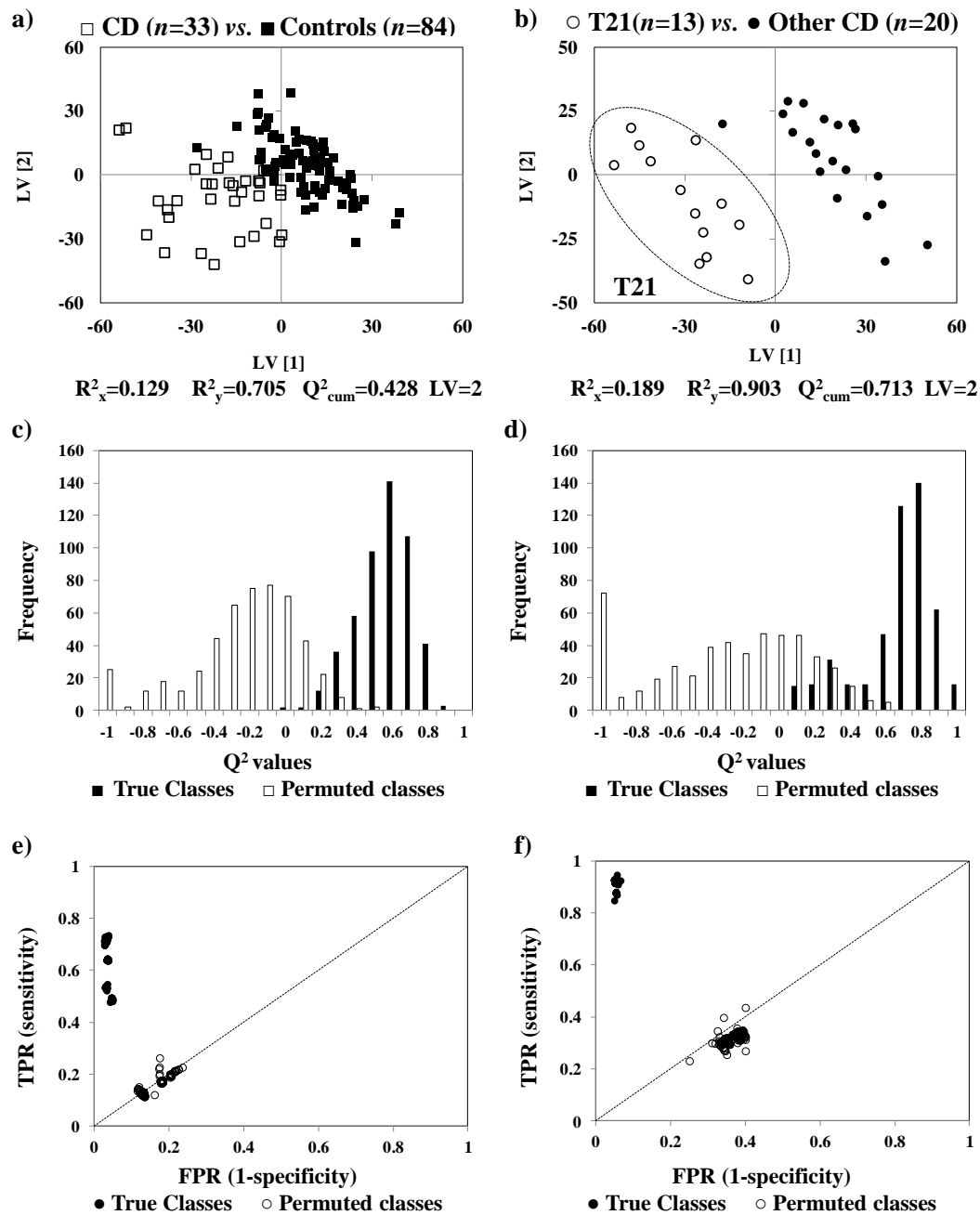


Figure 4.8: a) PLS-DA scores plot, c) Q^2 distributions plot and e) ROC plot of true and permuted classes, given by MCCV obtained for CD (□, $n=33$) vs. Controls (■, $n=84$) and b) PLS-DA scores plot, d) Q^2 distributions plot and f) ROC plot of true and permuted classes, given by MCCV obtained for T21 (○, $n=13$) vs. Other CD (●, $n=20$). TPR: true positive rate, FPR: false positive rate.

The PLS-DA model was found to have high predictive power (low overlapping of Q^2 distribution of true and permuted models) and MCCV parameters (CR 85%, sensitivity 87% and specificity 84%), these being summarized in Table A-III.2, Annex III and shown in Figure 4.8c,e. The analysis of the urinary profile explaining the difference between cases

and controls revealed 25 compounds, 4 unassigned (doublet 1.15 ppm, singlet 2.39 ppm, doublet 1.25 ppm and singlet at 4.54 ppm) and several spectral regions, these being listed in Table 4.2 left. The overall changes include higher excretion of amino acids (glutamine, histidine, isoleucine and threonine) with the exception of tyrosine that decreases, higher ketone bodies, glucose and pyruvate, as well as higher 2-KG, citrate, choline, creatine, creatinine, guanidoacetate (GAA), NMND and other metabolites related from the gut microflora.

The possibility to differentiate CD types was also attempted, this being only possible for T21 as a low number of samples were available for the remaining types (see Table 2.3 of Experimental Section). The PLS-DA model obtained after variable selection revealed a clear separation of T21 from other CD types, as shown in Figure 4.8b), with high model performance (CR 94%, sensitivity 92%, specificity 95%, summarized in Table A-III.2, Annex III) and predictive power (almost no overlapping of Q^2 obtained for true and permuted models), shown in Figure 4.8d,f). T21 cases were found to hold a distinctive NMR signature from other CDs, this being listed in Table 4.2 right, with specific variations noted for 9 metabolites (*N*-methyl-2-pyridone-5-carboxamide (2-Py), allantoin, glucose, glycine, histidine, lysine, malonate, NMND and taurine) and 10 unassigned resonances with variations specific of T21 noted with ^c in the same table. Out of these changes, only 2-Py, U10 (possibly related to diet/gut microflora, as was found to change due to non-fasting condition, Section 3.2) and 7 unassigned regions were found statistically different ($p < 0.05$).

The specific urinary metabolic signature of CD and the distinctive features of T21 were represented through the VIP-wheel. The resulting fingerprints of the disorders are shown in Figure 4.9, where the selected profile of both CD and T21 are shown in dark and light blue, respectively. This representation shows specific changes with high importance in CD, namely for pyruvate, citrate, 4-OH-hippurate, histidine and 1-methyl-histidine. For T21 cases, specific signatures were seen for unassigned regions 0.70-0.82, 1.40-1.85 and 2.21-2.23 ppm (these regions typically containing resonances from C-CH₃ and =C-CH₃).

Table 4.2: Metabolite/resonance changes in maternal urine of CD vs. controls and T21 vs. other CDs. ^a Chemical shifts selected by variables selection; TOCSY or STOCYSY correlated peaks in square brackets; s: singlet, d: doublet, t: triplet, q: quartet, dd: doublet of doublets, m: multiplet, br: broad. ^b Metabolites related to diet and/or gut microflora. ^c Changes specific of T21 disorders. ^d Consistent with previous findings published as result of this work (Diaz and Pinto *et al.*, 2011). Ui: unassigned compound i following numbering in Table 4.1. DMA: dimethylamine, GAA: guanidoacetate, 2-Py: *N*-methyl-2-pyridone-5-carboxamide, other abbreviations defined in Table 4.1. Only *p*-values<0.05 are shown.

CD n=33 vs. controls n=84			T21 n=13 vs. other CDs n=20		
Compound	δ_H /ppm and multiplicity ^a	Variation(effect size, <i>p</i> -value)	Compound	δ_H /ppm and multiplicity ^a	Variation(effect size, <i>p</i> -value)
1-Me-histidine ^b	3.72 s, 7.05 s, 7.78 s	$\uparrow(0.50, 3.2 \times 10^{-2})$	2-Py ^c	6.67 d, 8.33 d	$\downarrow(-0.73, 4.8 \times 10^{-2})$
2-KG	2.45 t, 3.01 t	$\uparrow(0.54, 3.0 \times 10^{-2})$	4-OH- hippurate ^b	6.98 d, 7.76 d	$\downarrow(-0.41)$
3-HBA	1.20, d	$\uparrow(0.46, 3.8 \times 10^{-2})$	Allantoin ^c	5.39 s	$\uparrow(0.43)$
4-DTA	1.23 d	$\downarrow(-0.16)$	Choline ^b	3.20 s	$\downarrow(-0.60)$
4-HPA ^b	6.86 d, 7.17 d	$\downarrow(-0.35)$	Glucose ^c	5.25 d, 3.23 dd	$\downarrow(-0.68)$
4-OH- hippurate ^b	3.95 s, 6.98 d	$\downarrow(-0.34, 2.3 \times 10^{-2})$	Glycine ^c	3.57 s	$\uparrow(0.39)$
Acetone	2.23 s	$\uparrow(0.43)$	Histidine ^c	4.01 q, 7.13 s, 7.98 s	$\downarrow(0.52)$
Acetoacetate	2.28 s, 3.45 s	$\uparrow(0.34)$	IS ^b	7.51 d, 7.70 d	$\downarrow(-0.47)$
Choline ^{b,d}	3.20 s	$\uparrow(0.35)$	Lysine ^c	1.73 m, 1.91 m	$\downarrow(-0.65)$
Citrate	2.54 d, 2.69 d	$\uparrow(0.45)$	Malonate ^c	3.11 s	$\downarrow(-0.49)$
Creatine	3.93 s	$\uparrow(0.40)$	NMND ^c	4.48 s, 8.90 d, 8.97 d	$\downarrow(-0.60)$
Creatinine	4.06 s	$\uparrow(0.37, 8.5 \times 10^{-3})$	Taurine ^c	3.43, t	$\downarrow(-0.40)$
DMA ^b	2.73 s	$\downarrow(-0.17)$	U1 ^c	0.63 s	$\uparrow(1.10)$
GAA	3.80 s	$\uparrow(0.56)$	U10 ^b	4.40 s, 8.78 d	$\uparrow(0.59, 1.1 \times 10^{-2})$
Glucose	5.25 d, 3.23 dd, 3.50 t, 3.44 m	$\uparrow(0.46)$	U16 ^c	1.25 d	$\downarrow(-0.26)$
Glutamine	2.47 m	$\uparrow(0.24)$	U20 ^c	2.01 s	$\uparrow(0.58)$
Histidine	4.01 q, 7.98 s	$\uparrow(0.37)$	U21 ^c	2.03 s	$\uparrow(0.25)$
Isoleucine	1.01 d	$\uparrow(0.41)$	U22 ^c	2.78 s	$\uparrow(0.70)$
Lactose	4.46 d	$\downarrow(-0.21)$	U23 ^c	2.84 s	$\uparrow(0.36)$
NMND	8.90 d, 8.97 d, 9.12s	$\uparrow(0.30)$	U24 ^c	3.30, s	$\downarrow(-0.52)$
Pyruvate	2.38 s	$\uparrow(0.28)$	U25 ^c	6.81 d	$\uparrow(0.57)$
Scyllo-inositol ^b	3.36 s	$\downarrow(-0.40)$	U26 ^c	7.01 s	$\downarrow(-0.75)$
Sucrose ^b	5.41 d	$\uparrow(0.54)$	U27 ^c	7.68 s	$\downarrow(-0.46)$
Threonine	4.26 dd	$\uparrow(0.27)$	Unassigned spectral regions		
Trigonelline ^b	4.44s, 8.84br, 9.12s	$\downarrow(-0.29)$	0.74-0.76		$\uparrow(0.81, 3.3 \times 10^{-2})$
Tyr	6.91 d	$\downarrow(-0.29)$	0.80-0.82		$\uparrow(1.30, 3.3 \times 10^{-3})$
U3	1.15 d [3.45, 3.55]	$\downarrow(-0.44)$	0.92-0.93		$\uparrow(0.92, 3.0 \times 10^{-2})$
U7	2.39 s	$\uparrow(0.63, 1.2 \times 10^{-2})$	1.40-1.43		$\uparrow(0.94, 2.0 \times 10^{-2})$
U16	1.25 d	$\downarrow(-0.21)$	1.65-1.68		$\uparrow(0.88, 1.6 \times 10^{-2})$
U19	4.54, s	$\downarrow(-0.21)$	1.78-1.81		$\uparrow(0.84, 2.4 \times 10^{-2})$
Unassigned spectral regions			1.84-1.85		$\uparrow(0.91, 2.4 \times 10^{-2})$
3.05-3.08		$\uparrow(0.38, 1.2 \times 10^{-2})$			
+ 6 selected regions, <i>p</i> >0.05, at 0-4.6 and 6-10 ppm					

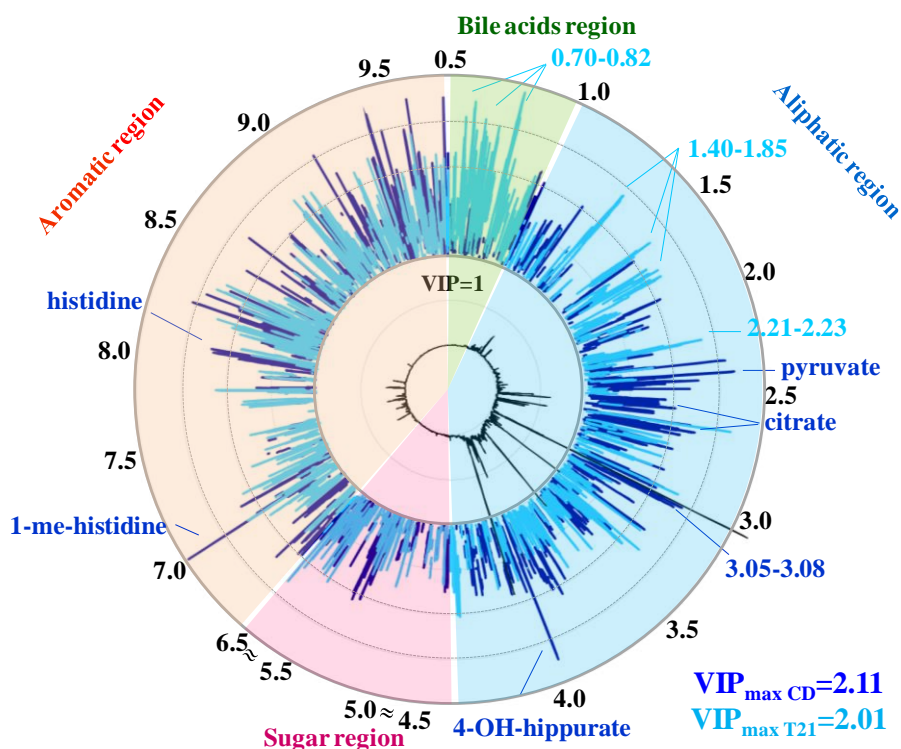


Figure 4.9: VIP-wheel representations of the NMR metabolite signatures obtained for CD (dark blue) overlaid with T21 (light blue). The average ^1H NMR spectrum of controls is represented in the inner circle, with the corresponding ppm scale shown in the outer black circle. Each dot represents a variable (spectral data point) selected by the intersection method, positioned in a radial VIP scale according to its VIP value. Some distinguishing features are indicated.

4.1.2.2 Proposed metabolic interpretation of the urinary signature of general CD and T21

A representation of the metabolic pathways affected in CD and T21 is shown in Figure 4.10, along with changes found for FM and CNS. In terms of the biochemical characteristics, CD cases were characterized by common effects to FM, namely regarding the higher excretion of ketone bodies and changes in nucleotide metabolism, along with evidence of enhanced TCA cycle activation (Table 4.2, Figure 4.10). Further common variations were found for unassigned singlet at 2.39 (increased) and the doublet at 1.15 ppm (decreased) in both CD and FM groups, reflecting some kind of common metabolic perturbation (unidentified at this stage). This finding hints at underlying general metabolic responses in cases of fetal pathologies, as CD are often associated with the occurrence of a malformed phenotype although with varying severity (see Table 1.5 in Introduction). Increased excretion of NMND suggests a perturbation in nucleotide metabolism,

particularly the tryptophan-NAD⁺ pathway (Salek *et al.*, 2007), as found for the FM group. Another perturbation found for CD is the higher excretion of citrate and 2-KG both intermediates of the TCA cycle and consistent with an enhancement of this cycle. Increased excretion the two latter metabolites were also found specifically in the CNS group but not in the heterogeneous FM group, thus suggesting a common perturbation between CD and CNS malformations. Increased ketone bodies (3-HBA, acetone and acetoacetate, the first being the only ketone body varying with individual significance) reflect perturbations in energy production, as was advanced for the FM group. Furthermore, higher 3-HBA has been reported in 1st trimester maternal blood serum of women carrying fetuses with T21 possibly associated with changes in brain growth and myelination as 3-HBA is an important substrate for the synthesis of phospholipids and sphingolipids (Bahado-Singh *et al.*, 2013).

A specific feature of CD relatively to FM, concerns the increased excretion of glucose and sucrose along with decreased lactose in CD (Figure 4.10), which is consistent with a derangement in sugar metabolism previously suggested by changes in AF (Graça, 2013). Moreover, increased glucose suggest its catabolic underuse, along with increased pyruvate, either due to metabolic changes and/or renal adaptations expected in healthy pregnancy progression (Creasy *et al.*, 2009), but not in relation to CD. Evidence of altered amino acids metabolism was also found in CD cases, namely by decreased tyrosine and increased glutamine, histidine, isoleucine and threonine along with decreased 4-DTA, a product of threonine catabolism (Appiah-Amponsah *et al.*, 2009) in CD cases. Changes in amino acids metabolism are supported by previous findings in AF (Amorini *et al.*, 2012, Graça, 2013). Higher GAA and creatine were also found in CD, possibly reflecting a perturbation creatine biosynthesis or a slowing down of urea cycle (Brosnan and Brosnan, 2010) although no relation has been previously reported regarding these pathway and CD. Finally, some relevant changes were also found for metabolites deriving from the gut microflora, namely for 4-HPA, 4-OH-hippurate, choline, DMA, *scyllo*-inositol and trigonelline. This finding unveils an important relationship between gut microflora and CD, although its exact nature cannot, at this point, be further explored.

In relation to the T21 group within the overall CD group, a distinctive profile was found relatively to other CDs (shown in light blue in Figure 4.9). The tandem decrease of 2-Py and NMND, both involved in nucleotide metabolism with 2-Py being formed from

NMND, suggest a pronounced perturbation in nucleotide metabolism. Several specific changes of T21 were found regarding amino acids excretion, namely increased glycine and decreased histidine, lysine and taurine in maternal urine. Changes in amino acid concentrations have been reported in AF of T21 cases namely decreased isoleucine, leucine, glycine, lysine, taurine, valine, ornithine and glutamate and increases in glutamine (Amorini *et al.*, 2012) with changes of glycine and glutamate possibly reflecting a negative effect of fetal neuronal functions. Furthermore, the variation found in this study for taurine gains additional value in relation to methionine metabolism. The gene encoding the enzyme cystathione beta synthase (CBS, enzyme catalyzing the conversion of homocysteine to cystathione) is located on chromosome 21, thus subjects with T21 have an enhanced activity of CBS leading to alterations in methionine metabolism (Obeid *et al.*, 2012). Another interesting finding is the increased excretion of allantoin, as found for the CNS group. This is consistent with the known role of oxidative stress in T21 (Bahado-Singh *et al.*, 2013). Indeed, higher allantoin levels have been previously reported in plasma of T21 subjects reflective of increased oxidative stress (Žitňanová *et al.*, 2004). An additional specific change of T21 was found for malonate. Malonate is a competitive inhibitor of succinate in the respiratory chain (Amaral *et al.*, 2012) and possibly reflects a perturbation at this level, although no further evidence supporting this possibility was found in literature.

With regard to the unassigned resonances, one last mention on the relevance of these peaks/compounds, as these play an important role in the definition of the T21 signature. This signature comprises 11 unassigned resonances, 10 of these being specific to the disease along with 7 spectral regions with individual significance, some of these being indicated in Figure 4.9. STOCYSY was performed in order to attempt improving the knowledge on possible metabolic relationships of the unassigned compounds. A relevant positive correlation was obtained for the singlet at 7.68 ppm (increased in other CDs), which correlated with creatinine (r 0.91). Still, no variation was found for creatinine or creatinine metabolism in this group, thus hampering the proposal of a relevant biochemical interpretation. Hence, a significant part of the metabolic signature of T21 is still unknown and important assignment work is still crucial so that a more complete metabolic picture can be drawn.

FETAL DISORDERS STUDY: AFFECTED METABOLIC PATHWAYS

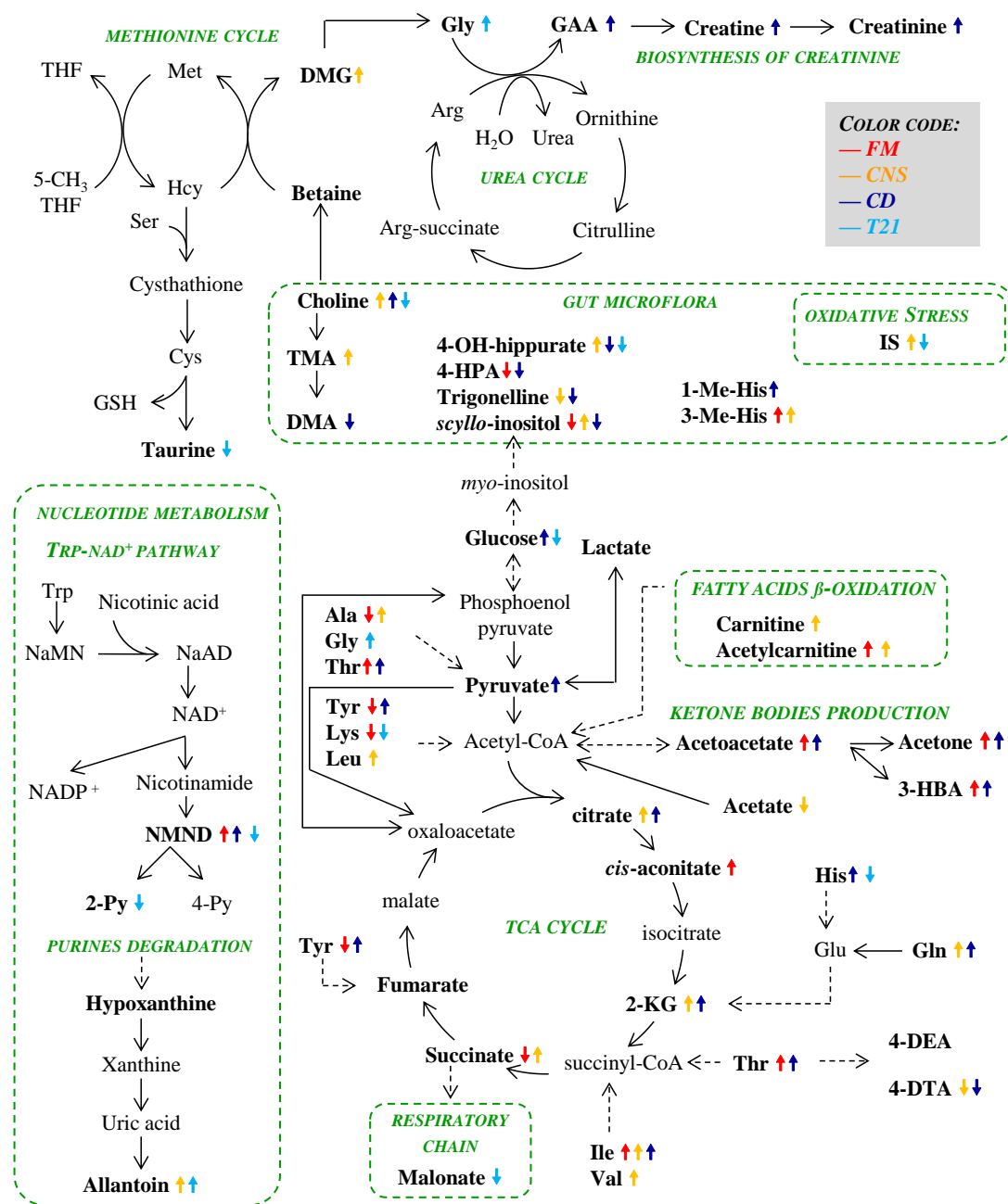


Figure 4.10: Schematic representation of the metabolic pathways possibly affected in FM, CNS, CD and T21 cases. Metabolites in bold were identified in ¹H NMR spectra of maternal urine. Colored arrows indicate direction of variation in each group. Dashed arrows indicate transformations occurring through several reactions. Amino acids in 3-letter code; NaMN: nicotinic acid mono nucleotide, NaAD: nicotinic acid adenine dinucleotide, NAD⁺: nicotinamide adenine dinucleotide, NADP: nicotinamide dinucleotide phosphate, ADP: adenosine diphosphate, NMND: *N*-methyl-nicotinamide, 2-Py: *N*-methyl-2-pyridone-5-carboxamide, 4-Py: *N*-methyl-4-pyridone-3-carboxamide, 4-DEA: 4-deoxyerythronic acid, 4-DTA: 4-deoxythreonic acid, 3-HBA: 3-hydroxybutyrate, DMA: dimethylamine, TMA: trimethylamine, DMG: dimethylglycine, GAA: guanidoacetate, Hcy: homocysteine, GSH: glutathione, THF: tetrahydrofolate.

4.2 Prediction of poor pregnancy outcomes through second trimester maternal urine metabolomics

4.2.1 Women later developing preterm delivery (PTD)

The possibility of predicting the development of PTD was attempted in this study by comparing control urine to maternal urine samples collected during the 2nd trimester, 11-20 weeks before delivery (sample group here named *pre*-PTD). A more detailed description on sample groups metadata is shown in Section 2.1.2 b), Table 2.2 of Experimental Section. For this sample group, the quest for novel biomarkers capable of predicting the disorder was carried out by comparing samples analyzed by ¹H NMR and UPLC-MS. Furthermore, external validation, using the initial NMR model used as a training set, was also attempted for 6 *pre*-PTD samples and 29 controls.

Table 4.3: Sample numbers, maternal age (in years), gestational age at sampling (in gestational weeks, g.w.) and interval from collection to diagnosis for controls and *pre*-PTD samples used in the ¹H NMR and UPLC-MS (right). Due to differences in sampling and acquisition two additional *pre*-PTD samples were available when UPLC-MS experiments were performed, the number of samples used in each model is not exactly the same.

Metadata	¹ H NMR work		UPLC-MS work	
	Control	<i>pre</i> -PTD	Control	<i>pre</i> -PTD
<i>n</i>	84	26	25	28
Maternal age (years)	21-42 (36)	26-41 (36)	26-39 (35)	21-41 (36)
Gestational age (g.w.)	15-26 (17)	16-21 (17)	15-26 (17)	16-21 (17)
g.w. to diagnosis	—	11-20 (17)	—	11-20 (17)

4.2.1.1 Unveiling a *pre*-diagnosis urinary metabolic signature of *pre*-PTD by NMR

The initial PLS-DA score plot (not shown) revealed a separation trend between *pre*-PTD and control samples, although insufficient results were obtained by MCCV (see Table A-III.2, Annex III). Thus, the spectra dataset was reduced by applying variable selection and the profile preceding the development of PTD was unveiled. The resulting PLS-DA scores plot showed a clear separation in the score plot, with satisfactory predictive power (Q^2 0.458), CR, sensitivity and specificity (84%, 79% and 84%, respectively). The obtained PLS-DA score plot along with Q^2 distribution and ROC plot, of true and permuted models, are shown in Figure 4.11.

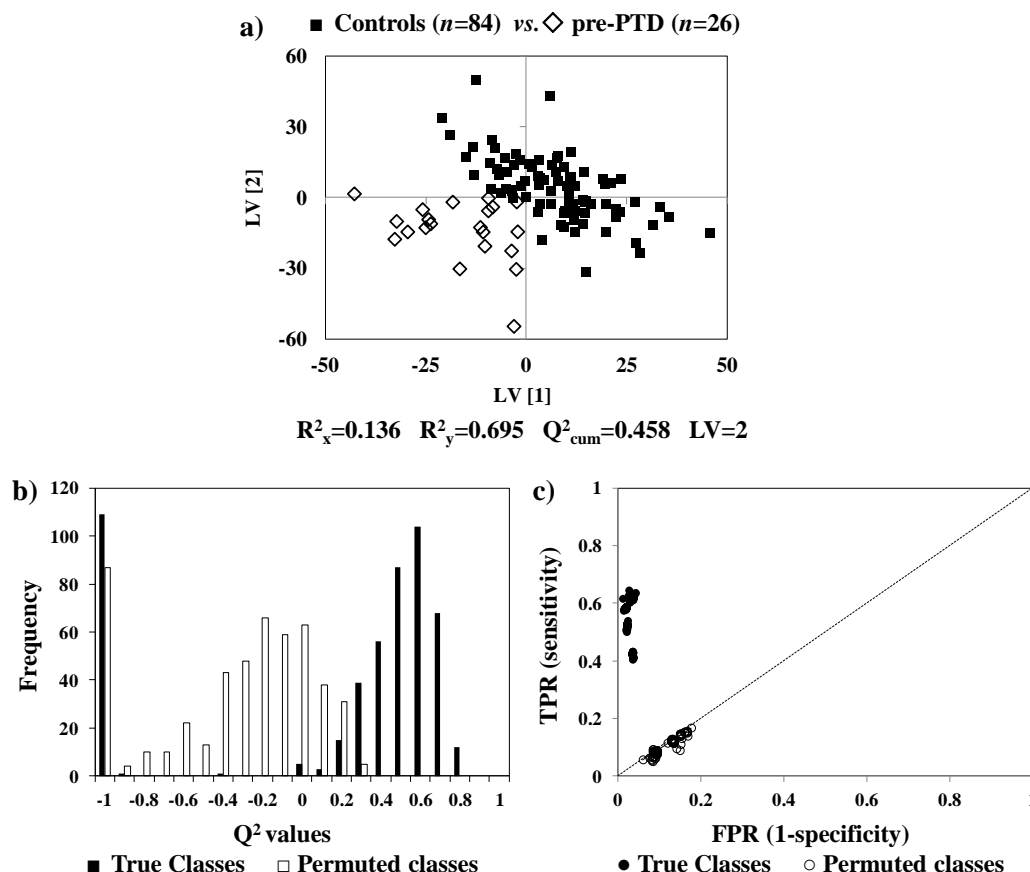


Figure 4.11: a) PLS-DA scores plot, b) Q^2 distributions plot and c) ROC plot of true and permuted classes, given by MCCV obtained for Controls (■, $n=84$) vs. *pre*-PTD (◇, $n=26$). TPR: true positive rate, FPR: false positive rate.

The metabolic signature here presented as potentially predictive of PTD, and explaining the separation trend seen in Figure 4.11a) is listed in Table 4.4. Variations were also evaluated relatively to other *pre*-diagnosis conditions evaluated in this work and changes considered specific are indicated with a ^c in Table 4.4 whilst changes common to all *pre*-diagnosis disorders (increased 4-DEA and decreased creatinine, possibly reflecting a general stress effect) are indicated with a ^d in the same table.

The profile of *pre*-PTD was analysed through a VIP-wheel representation of the selected variables and corresponding VIP value, as shown in Figure 4.12. Some distinctive findings of *pre*-PTD are indicated, namely for ascorbate, 4-OH-hippurate, unassigned singlet at 1.99 ppm (U4), 1.45-1.70 ppm and 6.73-6.76 ppm. The overall *pre*-PTD signature comprises changes in 21 metabolites, 4 unassigned compounds and 6 spectral regions (with $p<0.05$). Out of these changes only 3-methylhistidine and 4-HPA both

related with diet and/or gut microflora, decrease with individual significance in *pre*-PTD. Moreover increased 3-HBA, acetoacetate, GAA, isoleucine, threonine and unassigned singlet at 0.55 ppm, along with decreased ascorbate, betaine, histidine, and unassigned singlet at 4.54 ppm were found to be specific of *pre*-PTD.

Table 4.4: Metabolite/resonance changes in maternal urine for the *pre*-PTD group compared to controls. ^a Chemical shifts selected by variables selection; TOCSY or STOCYSY correlated peaks in square brackets; s: singlet, d: doublet, t: triplet, q: quartet, dd: doublet of doublets, m: multiplet, br: broad. ^b Metabolites related to diet and/or gut microflora. ^c Changes specific of the disorder, relatively to other pre-diagnosis disorders here evaluated. ^d Changes common to all pre-diagnosis disorders. ^e Consistent with an initial exploratory report published as result of this work (Diaz and Pinto *et al.*, 2011). Ui: unassigned compound i following numbering in Table 4.1 and Table 4.2. PAG: phenylacetylglutamine, 3-HIVA: 3-hydroxyisovalerate, 4-DEA: 4-deoxyerythronic acid, other abbreviations defined in Table 4.1 and Table 4.2. Only p-values<0.05 are shown.

<i>pre</i>-PTD n=26 vs. Controls n=84		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
2-HIBA ^e	1.36 s	↑(0.19)
2-Py	6.67 d, 8.33 d	↑(0.22)
3-HBA ^c	1.20 d	↑(0.44)
3-HIVA	1.27 s	↓(-0.13)
3-Me-histidine ^b	3.75 s, 7.15 s	↓(-0.42, 4.4×10 ⁻³)
4-DEA ^{d, e}	1.11 d	↑(0.46)
4-HPA ^b	6.86 d, 7.17 d	↓(-0.45, 1.0×10 ⁻²)
4-OH-hipurate ^b	6.98 d, 7.76 d	↑(0.41)
Ascorbate ^c	4.52 d	↓(-0.28)
Acetone	2.23 s	↑(0.68)
Acetoacetate ^c	2.28 s, 3.45 s	↑(0.51)
Betaine ^c	3.26 s	↓(-0.10)
Carnitine	3.23 s	↓(-0.23)
Citrate	2.54 d, 2.69 d	↑(0.38)
Creatinine ^d	3.05 s, 4.06 s	↓(-0.15)
GAA ^c	3.80 s	↑(0.26)
Histidine ^c	7.99 s	↓(-0.21)
Isoleucine ^c	1.01 d	↑(0.34)
Lactose	4.46 d	↓(-0.21)
PAG ^b	2.27 t, 4.18 m, 7.43 t	↓(-0.32)
Threonine ^c	4.26 dd	↑(0.45)
U4	1.99 s	↓(-0.43)
U28	2.96 q [2.88, 4.02]	↑(0.25)
U19 ^c	4.54 s	↓(-0.25)
U12 ^c	0.55 s	↑(0.35)
Unassigned spectral regions		
0.66-0.73		↑(0.51, 1.5×10 ⁻²)
1.37-1.44		↑(0.67, 1.4×10 ⁻⁴)
1.45-1.47		↑(0.66, 6.4×10 ⁻³)
1.60-1.65		↑(0.63, 3.2×10 ⁻³)
5.33-5.34		↓(-0.44, 2.3×10 ⁻²)
6.73-6.76		↓(-0.66, 3.9×10 ⁻³)
+ 13 selected regions, <i>p</i> >0.05		

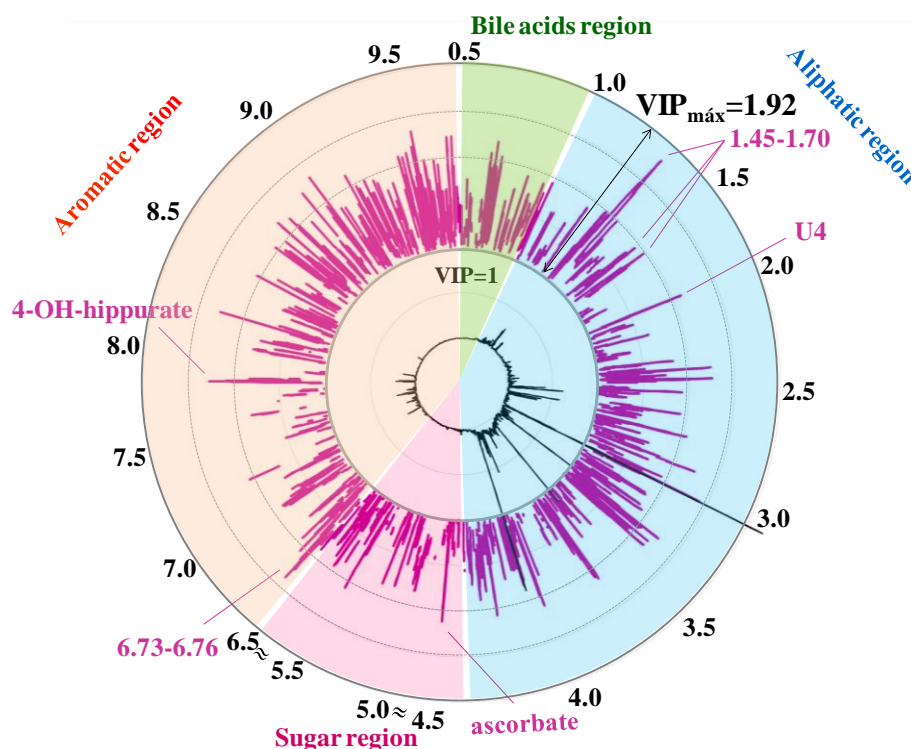


Figure 4.12: VIP-wheel representations of the NMR metabolite signatures obtained for *pre*-PTD. The average ^1H NMR spectrum of controls is represented in the inner circle, with the corresponding ppm scale shown in the outer black circle. Each dot represents a variable (spectral data point) selected by the intersection method, positioned in a radial VIP scale according to its VIP value. Some distinguishing features are indicated

4.2.1.2 Preliminary external validation of NMR models

External validation was attempted for *pre*-PTD by predicting new samples (29 controls and 6 *pre*-PTD samples) in the previously developed model. However, the same pitfall (discussed for the FM group) affected this prediction attempt as sample spectra had to be re-aligned, and PLS-DA repeated using only the set of variables previously selected. The resulting PLS-DA model has significantly decreased prediction power (Q^2 0.181) reflecting the absence of groups separation seen in the scores plot (Figure 4.13). Furthermore, unacceptable results were obtained for class prediction, shown in the confusion matrix, since no *pre*-PTD sample could be correctly classified, yielding a sensitivity of 0%.

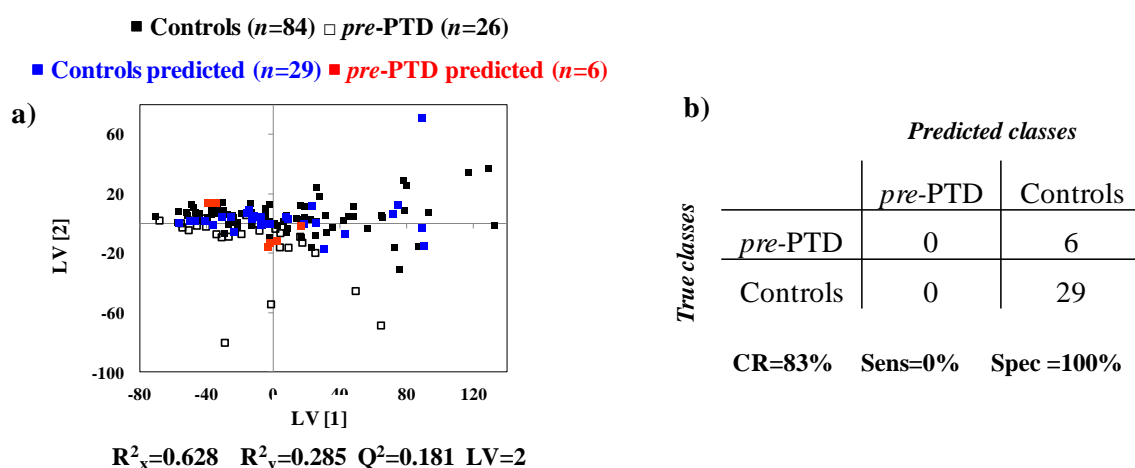


Figure 4.13: a) Original and predicted score plots and b) confusion matrices obtained for *pre*-PTD and controls CR: classification rate, Sens: sensitivity, Spec: specificity

These poor prediction results are probably due to a) the unbalanced number of samples used and b) data preprocessing steps, namely the realignment of the new NMR spectra. Thus, the use of groups balanced in number and more reproducible preprocessing steps such as bucketing is desirable, as discussed in section 4.1.1.3

4.2.1.3 Unveiling a *pre*-diagnosis urinary metabolic signature of PTD by UPLC-MS

To gain broader knowledge on the urinary NMR metabolic signature preceding PTD, maternal urine of *pre*-PTD and control groups were analyzed by UPLC-MS, as the method has higher sensitivity compared to NMR (pM of UPLC-MS vs. mM of NMR). Samples were run using both HSS and HILIC columns in both positive and negative ionization modes, as described in Section 2.3 of Experimental Section.

Figure 4.14 shows an example of a total ion chromatogram (TIC) of control maternal urine sample, obtained using the HSS column in the positive ionization mode (ESI+). As done for the NMR data, these complex datasets were converted into tables and further subjected to MVA in order to extract class-related changes. However, prior to MVA, important processing steps must be performed in order for relevant and reliable information to be obtained. Two processing methods were tested and results are described as follows:

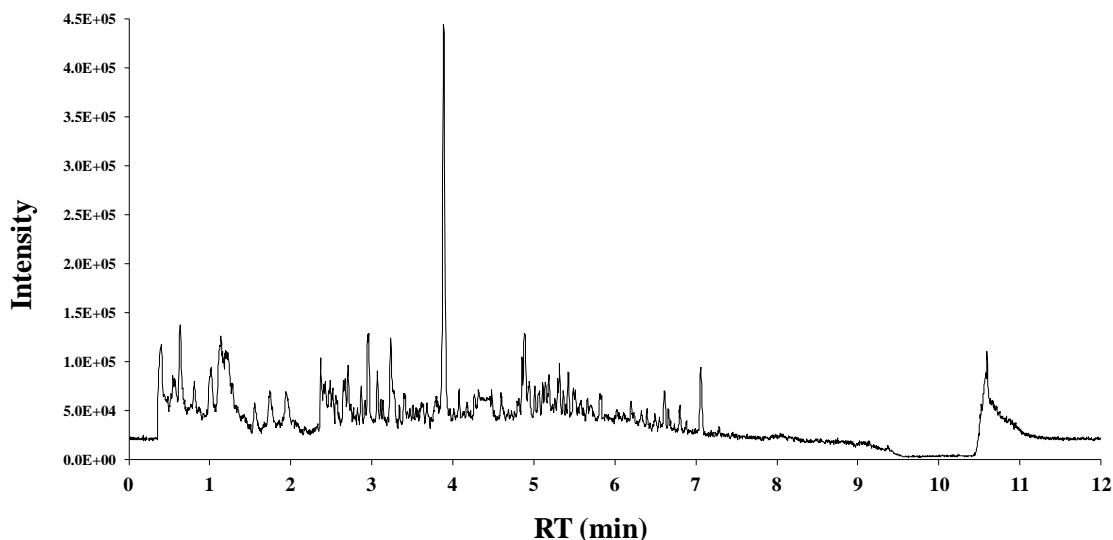


Figure 4.14: Total ion chromatogram (TIC) of a control urine sample using HSS column in ESI+ mode.

Results obtained with processing method 1

As, in the context of this thesis, a variable selection method based on multivariate parameters was developed and found extremely powerful for the unveiling of subtle disease profiles analyzed by ^1H NMR, the first approach was to apply the same methodology to the UPLC-MS data. Briefly, UPLC-MS data was processed (features were grouped, aligned and normalized), PLS-DA was computed and variables (in this case, features, i.e. a pair $\text{RT}_{\text{m/z}}$) were selected using the intersection method (i.e. simultaneously $\text{VIP} > 1$, $\text{VIP}/\text{VIP}_{\text{cvSE}} > 1$ and $|\text{b}_{\text{coeff}}/\text{b}_{\text{coeffcvSE}}| > 1$, a more detailed description of the method having been presented in Section 4.1.1.1).

The first approach was to calculate the PLS-DA of controls *vs.* *pre*-PTD of each dataset (HSS/ESI+, HSS/ESI-, HILIC/ESI+, HILIC/ESI-) using UV and Pareto scaling and Log transformation. The robustness and predictive power of all PLS-DA models was evaluated by MCCV and, for all models, insufficient results were obtained (high overlapping of Q^2 values of true and permuted models, low CR, specificity and sensitivity). Thus, as was done for NMR data, each dataset (Column/Detection mode/Scaling or transformation) was subjected to variable selection and the resulting matrix resubmitted to PLS-DA and MCCV. Apparently good and reliable separations, between controls and *pre*-PTD samples, were found, further supported by the resulting model Q^2 and validation

parameters (Table A-IV, Annex IV). Features were selected for further analysis based on the *S*-plot (i.e. covariance vs. correlation of each feature and the 1st LV) and features with high covariance ($|p| > 0.02$) and correlation ($|p_{\text{corr}}| > 0.6$). Furthermore, these features were evaluated in terms of their covariance and covariance standard error ratio ($p[1]/p_{\text{cvSE}}[1]$) with those having a ratio higher than 4 (i.e. error below 25%) being selected for univariate analysis (*p*-value and effect size).

The PLS-DA model obtained using HSS/ESI+/Log transformation yielded a CR of 97%, sensitivity 96% and specificity 97%. Nevertheless, only one feature (RT= 2.92 min and *m/z* 590.04) was found to fulfill all the established requirements (described in the above paragraph). However, inspection of the original dataset revealed that this same feature was in fact noise. This drawback was found for all features selected through PLS-DA models of datasets obtained with this variable selection methodology. Thus, it was concluded that the method is not suitable for UPLC-MS data processing and the second processing method was done with results presented in the following section.

Results obtained with processing method 2

The second processing method employed involves features selection based on three univariate steps, comprising an initial prefilter step, a MinFrac filter and a coefficient of variation filter. The prefilter was applied before grouping and alignment, which removed features that are not detected in at least 5 scans with intensity higher than 1000. This prefilter step is extremely useful in removing noise and background from the initial dataset, subsequently reducing their influence of the definition of features. Secondly, features were grouped, aligned and normalized, followed by the application of a MinFrac Filter, meaning that features not present in at least 50% of at least one of the groups are removed from the dataset. Finally, a coefficient of variation filter is also applied, with features with coefficient of variation higher than 30% (calculated for the QCs) were removed.

The initial appreciation of the resulting datasets obtained with this processing method and applied to each pair column/detection model/scaling or transformation revealed very good reproducibility of the UPLC runs. PCA score plots of samples (controls and *pre*-PTD) and QC show clustering of the QC samples for each dataset (Column/Detection mode/Scaling or transformation), these being shown in Figure A-IV.1, Annex IV.

Controls and *pre*-PTD samples, analyzed by HSS/ESI+, HSS/ESI-, HILI/ESI+ and HILIC/ESI-, were compared using UV and Pareto scaling and Log transformation, the resulting PLS-DA and MCCV parameters being summarized in Table A-IV.2, Annex IV. The best result was obtained for HSS/ESI+/Log, showing a separation trend between classes in the PLS-DA score plot and marginally satisfactory prediction power, seen through the low overlapping of Q^2 of true and permuted models and robustness, with CR of 69%, sensitivity 69% and specificity 70%, as shown in Figure 4.15. However, two control outliers were identified, these being noted in the PLS-DA score plot shown in Figure 4.15a, and the removing of these two samples caused the dropping of Q^2 to -0.169, suggestive an unstable model and very dependent of these two samples.

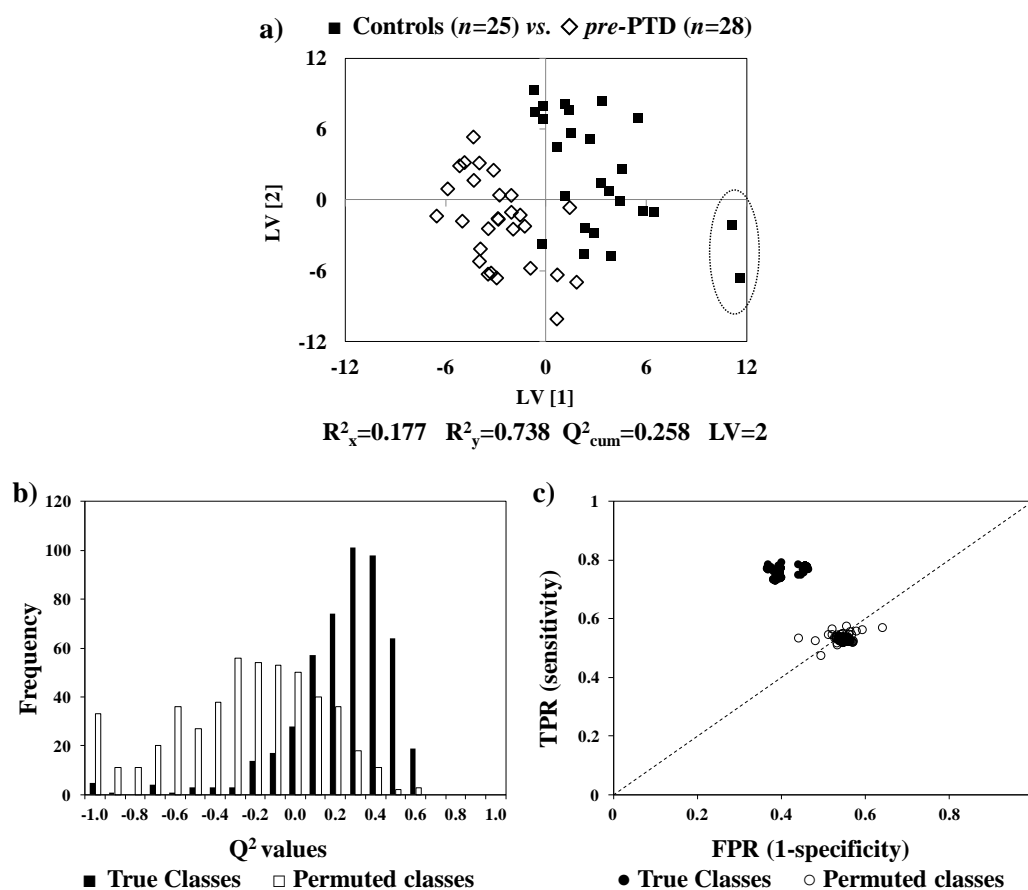


Figure 4.15: a) PLS-DA scores plot, b) Q^2 distributions plot and c) ROC plot of true and permuted classes, given by MCCV obtained for Controls (■, $n=25$) vs. *pre*-PTD (◇, $n=28$). TPR: true positive rate, FPR: false positive rate.

In spite of the results obtained through MVA, two features fulfilling the pre-established requirements (i.e. high covariance ($|p| > 0.02$) and correlation ($|p_{\text{corr}}| > 0.6$, and covariance to covariance standard error ratio ($p[1]/p_{\text{cvSE}}[1] > 4$) were found and selected for further analysis (RT 3.22 min, m/z 673.14 and m/z 689.11), after removing the two control outliers. The MS spectrum collected at RT 3.22 (Figure 4.16a), shows that these two features (indicated with a red arrow) are well defined, not noise or background as those found with the previous method.

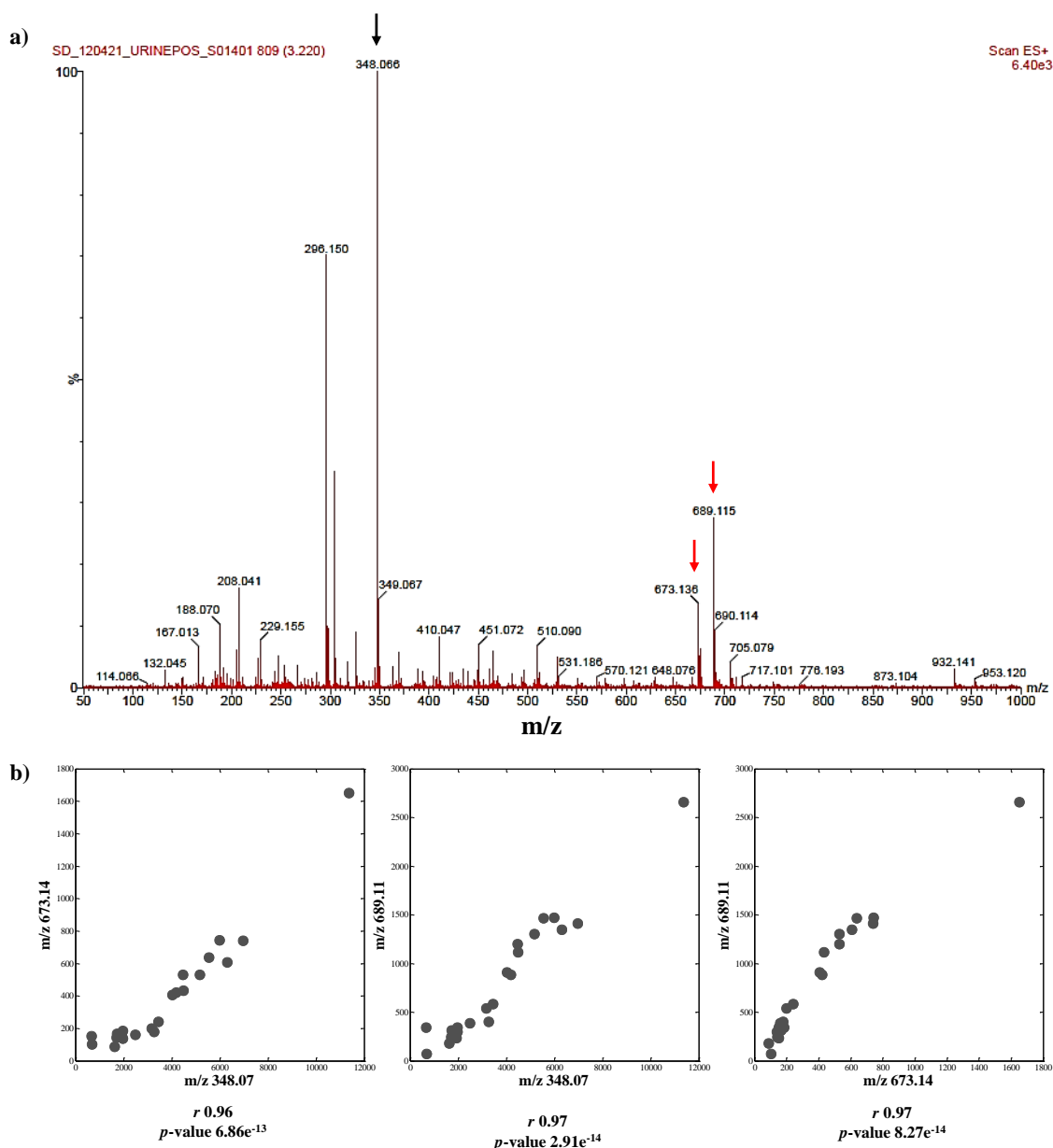


Figure 4.16: a) MS spectrum, of a control sample, recorded at RT 3.22 min. Red arrows indicates two features selected through MVA and black arrow indicates a possible m/z belonging to the same compound. b) Scatter plots of feature intensities, with corresponding r and p -value of the correlation.

Furthermore, the m/z 348.07 was found eluting at the same RT but with higher intensity than m/z 673.14 and 389.11, possibly suggesting that it belongs to the same compound. To confirm this, features intensities were correlated in the control group, as these were increased in this class (with exception of one control sample with high intensity of m/z 348.07 but not the other two features). The resulting scatter plots are shown in Figure 4.16b) along with the r and p -value of the correlation between each pair of features, where it is shown that the three features probably belong to the same compound, as supported by the high r ($r < 0.96$) and low p -value (10^{-3} - 10^{-4} orders of magnitude). Thus, univariate comparison of the intensities of these three features in the control and *pre*-PTD samples (without the three outlier samples) was done, the resulting effect size and p -value being shown in Table 4.5.

Table 4.5: Changes in maternal urine for *pre*-PTD compared to controls analyzed by HSS/ESI⁺. Only p -values < 0.05 are shown.

<i>pre</i> -PTD ($n=28$) vs. Controls ($n=22$)		
RT(min)	m/z	Variation (effect size, p -value)
3.22	689.11	↓ (-0.75, 5.65×10^{-3})
3.22	673.14	↓ (-0.67, 1.66×10^{-2})
3.22	348.07	↓ (-0.36)

The results obtained with univariate comparison and correlations suggest that the three features belong to the same compound, those with m/z 673.14 and 689.11 probably being dimers of the m/z 348.07. At this stage, however, no possible assignment was found, hindering any metabolic interpretation. Thus, Statistical Heterospectroscopy (SHY) was attempted to possibly unveil these features assignment or a common metabolic relationship.

4.2.1.4 Statistical correlation of NMR and MS data

SHY is a powerful statistical tool to aid metabolite assignment and to unveil metabolic relationships, as presented in previous reports (Crockford *et al.*, 2006, Coen *et al.*, 2007, Wang *et al.*, 2008, Graça *et al.*, 2012). The concept beyond SHY is similar to that of 1D STOCSY (presented in Section 3.1), i.e., NMR peaks and MS features

corresponding to the same molecule have high covariance and correlation, enabling the assignment of a given compound detected by both analytical techniques. Furthermore, metabolically related compounds also show meaningful covariance and correlation enabling the identification of a shared metabolic relationship.

SHY of MS and NMR data was initially tested for some features of known assignment, as exemplified in the 1D STOCY shown in Figure 4.17a). In this example the MS feature corresponding the trigonelline ion $M+H^+$ (detected with the HSS column in the positive ionization mode, eluting at 4.44 min with m/z 139.06) was correlated with the NMR spectra. A high correlation and covariance was found between the NMR spin system of trigonelline (singlet 4.44 ppm, multiplet 8.78 ppm and singlet 9.12 ppm) and the MS feature, with r 0.98 and p -value 3.18×10^{-22} . This correlation was confirmed by scatter plotting the intensities of both NMR and MS dimensions.

MS features found relevant for the separation of *pre*-PTD and control samples were also correlated with the NMR data and, for the feature with m/z 673.14, the obtained 1D STOCY is shown in Figure 4.17b). In spite of an apparently good correlation between the trimethylamine-*N*-oxide (TMAO) resonance, with r 0.73 and p -value 1.17×10^{-4} , and feature with m/z 673.14, the scatter plot of the intensities in the MS and NMR dimensions showed that this was a spurious correlation supported by only one sample (noted with an arrow in Figure 4.17b). SHY was repeated for the remaining two MS features, the correlation result being similar to that shown in the previous example.

Results obtained for the *pre*-PTD group analyzed by UPLC-MS showed significant changes for only two features, although no assignment was obtained hindering any further metabolic interpretation. Nevertheless, relevant conclusion can be drawn regarding UPLC-MS processing, namely the inadequacy of the VIP- and b-coefficient-based variable selection methodology application to the complete UPLC-MS dataset, as noise and background variables were being selected and wrongly yielding apparently reliable models. Furthermore, the importance of applying noise and background filters in this type of datasets is reinforced, in order to avoid erroneous conclusions and the suggestion of supposedly disease markers. Future works entails the analysis of other disorders groups, namely *pre*-IUGR, *pre*-PE and GDM, for which data has been acquired although results were not available in time for inclusion in this thesis.

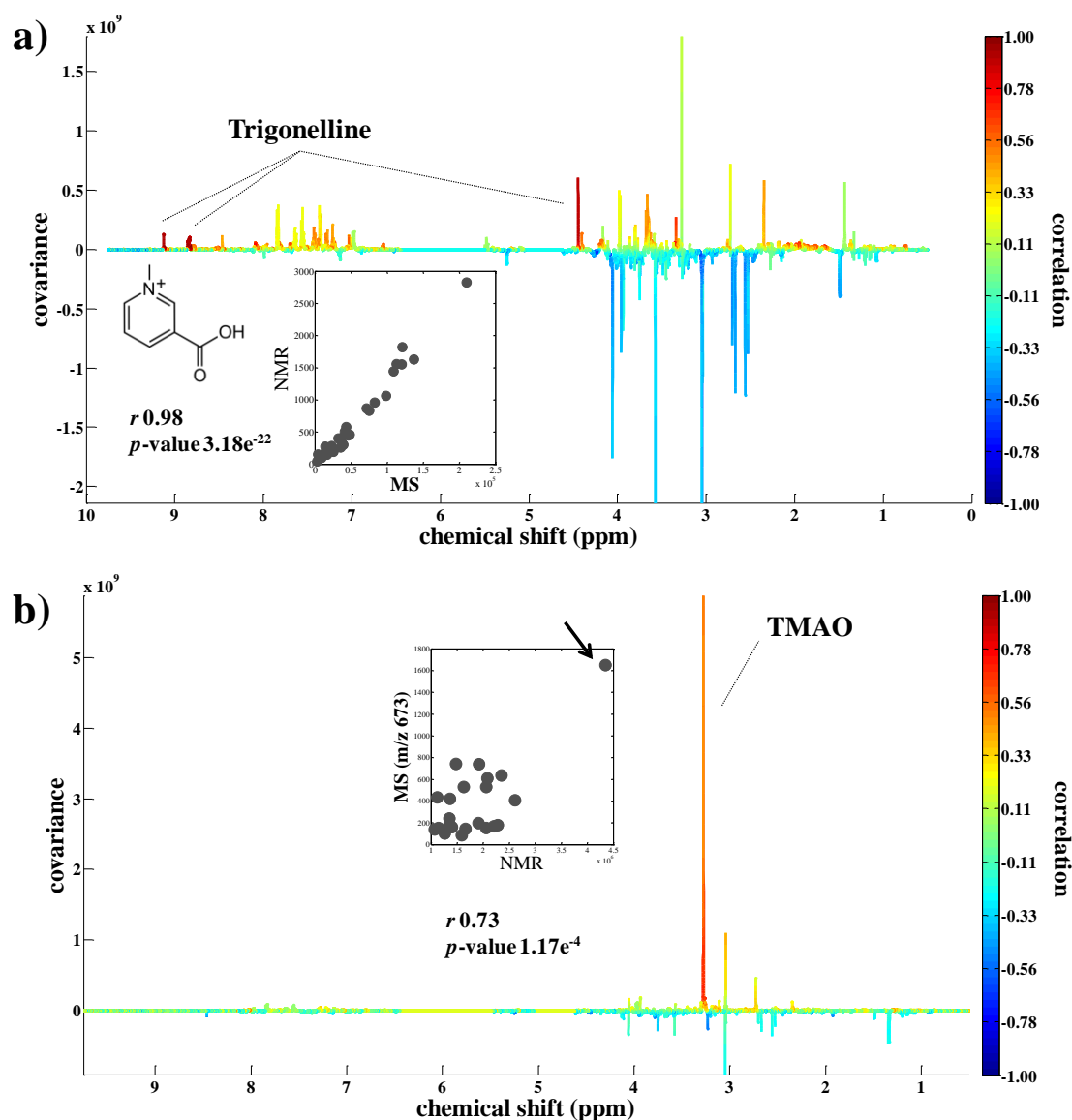


Figure 4.17: MS/NMR SHY analysis of a) trigonelline MS feature (139.06 m/z, 4.44 min) and b) relevant MS feature (673.14 m/z, 3.22 min) and the NMR spectra of urine. Arrow notes one sample compromising the correlation result.

4.2.1.5 Proposed metabolic interpretation of *pre*-diagnosis urinary changes in PTD

In spite of the premise that broader and complementary knowledge could be gained by analyzing *pre*-PTD samples through both NMR and UPLC-MS no significant findings could be obtained with the later technique, which shows the crucial importance of factors such as analysis conditions choice (e.g. LC column, MS detection method), analytical

reproducibility, increased dataset complexity due to the enhanced sensitivity and higher difficulty in compounds assignment. Hence, a possible metabolic interpretation could only be advanced based on the NMR data, albeit being less sensitive. A general scheme of the metabolic pathways found, in this study, to be affected in *pre*-PTD is depicted in Figure 4.22, along with perturbations found for other *pre*-diagnosed conditions which will be discussed in the following sections.

The overall changes noted in ^1H NMR spectra of urine from women who further developed PTD fit previously advanced hypothesis whilst also unveiling new information. Increased excretion of ketone bodies acetone, acetoacetate and 3-HBA, the latter two being specific of *pre*-PTD were found in this work, consistently with reported changes in AF (Graça, 2013). The increased ketone bodies observed here with the tandem decrease of carnitine suggest enhanced lipid oxidation several weeks before delivery. In fact, decreased lipids in umbilical cord blood plasma of preterm babies and increased lipids in their mothers blood plasma have been found, at the time of delivery, reflecting increased oxidation in the fetal compartment or altered placental transfer (Tea *et al.*, 2012). Decreased ascorbate was also found specific of *pre*-PTD, possibly in connection with oxidative stress. Ascorbate directly strengthens and stimulates collagen synthesis, and protects fetal membranes from ROS damage (a possible cause of membranes rupture) (Woods Jr *et al.*, 2001). Moreover, the hypothesis of increased oxidative stress during 2nd trimester of women later developing PTD has been advanced and supported by increased allantoin contents in AF. Still, in this work no changes were found for this metabolite in maternal urine. The *pre*-PTD group was also characterized by increased threonine and isoleucine, along with decreased histidine excretion. Increased threonine and decreased isoleucine and histidine has been found in AF possibly due early perturbation of the fetal-placental unit altering amino acid transfer or changes in protein synthesis (Graça *et al.*, 2012, Graça, 2013). Other specific changes registered in *pre*-PTD include decreased betaine and increased GAA, these metabolites involved in methionine metabolism and urea cycle, respectively. Betaine donates a methyl group to homocysteine to form methionine, also generating dimethylglycine (DMG). Previous reports have indicated hyperhomocysteinemia (increased homocysteine in blood) at 24-37 g.w. of women at risk of PTD and perturbations in the methionine pathway have been implicated in several adverse pregnancy outcomes (Dasarathy *et al.*, 2010, Micle *et al.*, 2012). Although no

concomitant changes were found in this work for other metabolites involved in methionine metabolism (i.e. choline or DMG), decreased betaine could indicate an early perturbation of the methionine cycle operation. The increased GAA urinary excretion may suggest a slowing down of the urea cycle (Brosnan and Brosnan, 2010) although no reference supporting this hypothesis was found in other reports. Relevant variations were also found in metabolites related to the gut microflora, namely for increased 4-OH-hippurate and decreased PAG, 3-Me-histidine and 4-HPA, the former two being the only metabolites with individual significance ($p < 0.05$) in this profile. These results suggest an important and novel role of gut microflora in the PTD pathological process, although the nature of the relationship remains, at this stage, unknown.

4.2.2 Women later developing preeclampsia (PE)

The possibility of predicting PE through urine metabolomics was also attempted by comparing samples collected 14-21 weeks before PE diagnosis and controls. A more complete description of samples metadata and diagnosis criteria is shown in Section 2.1.2 of the Experimental Chapter.

4.2.2.1 Unveiling a *pre*-diagnosis urinary signature of PE

In this work, a limited number of samples of women further developing PE (*pre*-PE, $n=9$) was obtained and compared to the control group ($n=84$). An initial PLS-DA revealed no significant changes in the ^1H NMR profile when considering the original dataset, as supported by the MCCV results in Table A-III.2, Annex III. However, when reducing the dataset with the variable selection method previously presented, a significant improvement was observed in the PLS-DA model and validation, with satisfactory predictive power given by Q^2 of true and permuted models, CR 94%, sensitivity 75% and specificity 94% as shown in Figure 4.18 and Table A-III.2, Annex III.

The analysis of the variables selected, thus explaining the separation visible in the PLS-DA score plots in Figure 4.18 revealed a metabolic signature potentially predictive of PE, this being listed in Table 4.6. *Pre*-PE samples were characterized by changes

21 metabolites, 4 unassigned spin systems (singlet at 0.93 ppm, 1.99 ppm, 4.40 ppm and doublet at 8.78 ppm, and singlet at 4.51 ppm) and 2 regions with $p < 0.05$. These changes were analyzed in terms of individual significance and effect size, along with specificity, relatively to other *pre*-diagnosis disorders, these being indicated with a ^c whilst changes common to all *pre*-diagnosis disorders (possibly reflecting a general stress effect) are indicated with a ^d in Table 4.6. The *pre*-PE metabolic signature was plotted by means of a VIP-wheel, this representation enabling the visualization of the selected variables with higher VIP value. The obtained plot is shown in Figure 4.19 with indication of some distinguishing features.

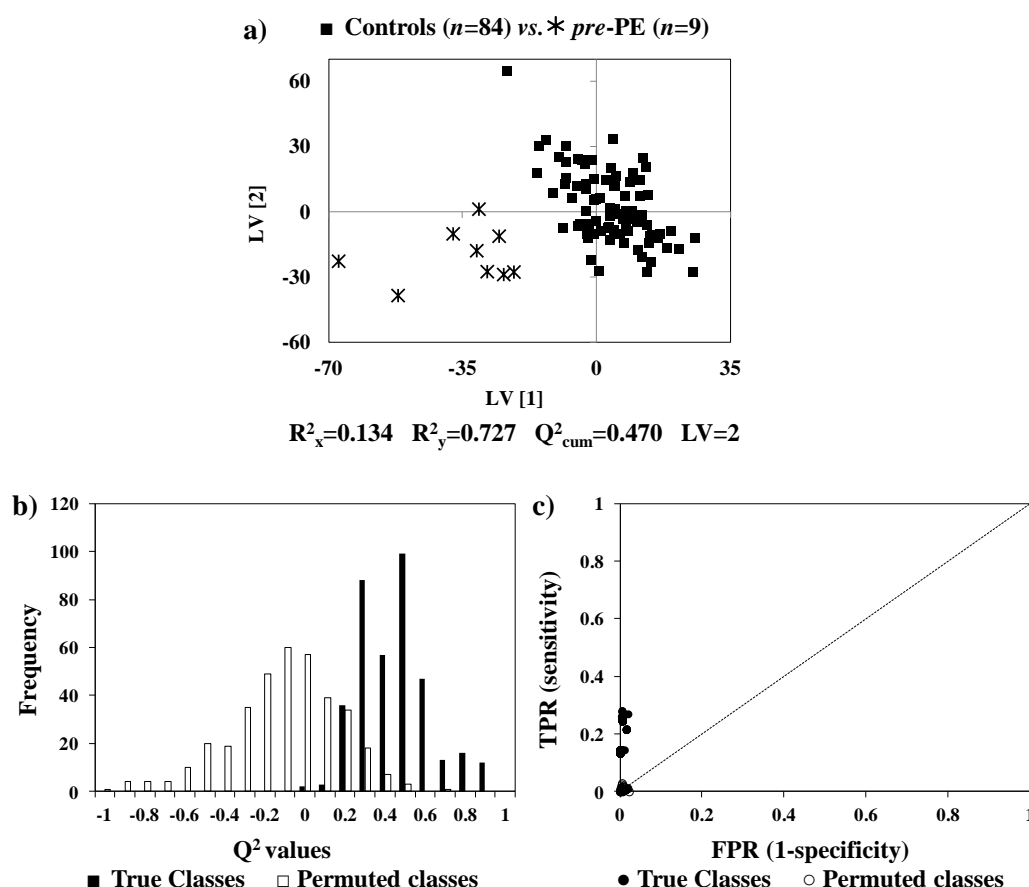


Figure 4.18: a) PLS-DA scores plot, b) Q^2 distributions plot and c) ROC plot of true and permuted classes, given by MCCV obtained for *pre*-PE (*, $n=9$) vs. Controls (■, $n=84$).

Table 4.6: Metabolite/resonance changes in maternal urine for *pre*-PE compared to controls.

^aChemical shifts selected by variables selection; s: singlet, d: doublet, t: triplet, q: quartet, dd: doublet of doublets, m: multiplet, br: broad. ^b Metabolites related to diet and/or gut microflora.

^cChanges specific of the disorder, relatively to other *pre*-diagnosis disorders here evaluated.

^dChanges common to all *pre*-diagnosis disorders. Ui: unassigned compound i following numbering in Table 4.1, Table 4.2 and Table 4.4 *p*-CS: *p*-cresol sulphate, other abbreviations defined in Table 4.1, Table 4.2 and Table 4.4. Only *p*-values < 0.05 are shown.

<i>pre</i>-PE <i>n</i>=9 vs. Controls <i>n</i>=84		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
2-KG	2.45 t, 3.01 t	↓(-0.49)
2-Py	8.33 d	↑(0.62)
3-Me-histidine ^b	7.15 s	↑(0.33)
4-DEA ^d	1.11 d, 4.12 m	↑(1.25)
4-OH- hippurate ^b	7.76 d	↓(-0.03)
Acetate ^c	1.92 s	↓(-0.26, 2.9×10 ⁻²)
Carnitine	3.23 s	↓(-0.45)
<i>Cis</i> -aconitate	3.12 d	↑(0.10)
Citrate	2.54 d, 2.69 d	↑(0.31)
Creatinine ^d	3.05 s, 4.06s	↓(-0.18)
Formate ^c	8.46 s	↓(-0.21)
Fumarate ^c	6.53 s	↓(-0.20)
Galactose	5.28 d	↓(-0.47, 2.0×10 ⁻²)
Glutamine ^c	2.15 m, 2.47 m	↑(0.05)
Hippurate ^b	3.97 d	↑(0.34)
Isoleucine ^c	1.01 d	↓(-0.26)
IS ^{b, e}	7.51 d, 7.70 d	↑(0.52)
Lactose	4.46 d	↓(-0.43)
NMND	4.48 s, 8.90 d, 8.97, d,	↑(0.22)
PAG ^b	2.27 t, 3.67 d	↓(-0.19)
<i>p</i> -CS ^b	2.34 s, 7.21 d, 7.28 d	↓(-0.34)
<i>Scyllo</i> -inositol ^b	3.36 s	↓(-0.68, 1.3×10 ⁻²)
Succinate	2.41 s	↓(-0.38, 1.5×10 ⁻²)
Sucrose	5.41 d	↑(0.71)
Trigonelline ^b	4.44 s, 8.84 br,	↓(-0.10)
	9.12 s	
Tyrosine	6.91 d	↓(-0.33)
U14	0.93 s	↓(-0.10)
U4	1.99 s	↓(-0.04)
U10 ^b	4.40 s, 8.78 d	↓(-0.06)
U35 ^c	4.51 s	↑(0.33)
Unassigned spectral regions		
1.28-1.30		↓(-0.30, 4.5×10 ⁻²)
1.54-1.58 ^e		↑(0.05, 4.8×10 ⁻²)
+ 28 selected regions, <i>p</i> >0.05, at 0-4.6 and 6-10 ppm		

In spite of the exploratory nature of this model (due to the limited number of samples for the *pre*-PE group), the results obtained revealed some apparently specific changes of the disease, these although requiring future confirmation in a larger cohort.

Moreover, the MCCV confirmation of the PLS-DA model reliability supported the possibility of a possible metabolic interpretation to be advance.

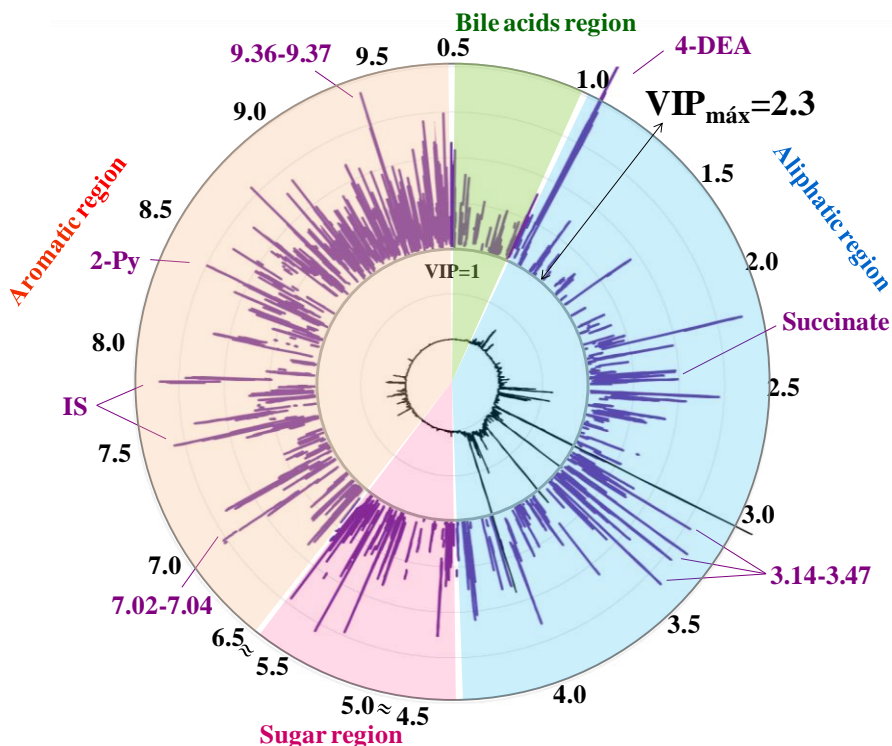


Figure 4.19: VIP-wheel representations of the NMR metabolite signatures obtained for *pre*-PE. The average ^1H NMR spectrum of controls is represented in the inner circle, with the corresponding ppm scale shown in the outer black circle. Each dot represents a variable (spectral data point) selected by the intersection method, positioned in a radial VIP scale according to its VIP value. Some distinguishing features are indicated.

4.2.2.2 Proposed metabolic interpretation of *pre*-diagnosis urinary changes in PE

In this cohort, apparently specific *pre*-PE variations were found for acetate, formate, fumarate, isoleucine and glutamine (although with a residual change), IS and an unassigned singlet at 4.51 ppm. An interesting finding is the concomitant change in TCA cycle precursor (decreased acetate) and intermediates (decreased fumarate and succinate along with increased *cis*-aconitate and citrate), suggestive of a profound perturbation in this pathway (Figure 4.22). Increased citrate and decreased succinate had been found in AF of women diagnosed with PE (Bock, 1994). Moreover, decreased succinate and acetate have been reported in 1st trimester maternal plasma of women latter developing PE, the

latter possibly due to changes in lipids metabolism (Bahado-Singh *et al.*, 2012). Another finding consistent with changes in lipids metabolism is the carnitine decrease found in this cohort, which is suggestive of increased lipids oxidation (Cho and Cha, 2005). Carnitine had been previously found to increase in 1st T maternal serum of later developing PE women, as a reflection of enhanced metabolism of lipids (Bahado-Singh *et al.*, 2012). Another specific change of *pre*-PE, within the *pre*-diagnostic disorders here studied, is the increased urinary excretion of IS. Increased urinary IS can reflect increased oxidative stress, as suggested for the FM study, as this metabolite has been found to mediate this process (Dou *et al.*, 2007). Moreover, increased oxidative stress has been suggested by changes in placental extracts of PE pregnancies (Dunn *et al.*, 2012). Nucleotide metabolism appears to be also altered, as suggested by the tandem increase in NMND and 2-Py found in this cohort. Increased urinary excretion of NMND and 2-Py suggest alterations in nucleotide metabolism and has been reported in connection to type 2 diabetes (Salek *et al.*, 2007), but not to PE. Regarding amino acids, the *pre*-PE group was associated with decreased isoleucine and tyrosine, the first having been previously found to increase in maternal urine (Glew *et al.*, 2004) and the second found to increase in maternal plasma (Turner *et al.*, 2008), both at the *post*-diagnosis state.

4.2.3 Women carrying fetuses later diagnosed with intrauterine growth restriction (IUGR)

For this study maternal urine of women who were diagnosed with intrauterine growth restriction (*pre*-IUGR) of the fetus later in pregnancy was compared to controls. A detailed description of samples metadata and diagnosis criteria is shown in Section 2.1.2 of the Experimental Section.

4.2.3.1 Unveiling a *pre*-diagnosis urinary metabolic signature of IUGR

As for *pre*-PE, only a limited number of samples were available for the study of *pre*-IUGR ($n=10$), hindering a more definite interpretation of the findings. These samples were compared to healthy controls ($n=84$) and, in spite of a separation trend being visible in the PLS-DA score plot (not shown), the MCCV revealed insufficient results, as shown

in Table A-III.2, Annex III. Thus, the original dataset was compressed, by applying the variable selection methodology here developed. The resulting PLS-DA model confirmed the separation between classes and MCCV validated the models' robustness (CR 94%, sensitivity 99% and specificity 94%) and predictive power of the model, as shown in Figure 4.20 and Table A-III.2, Annex III.

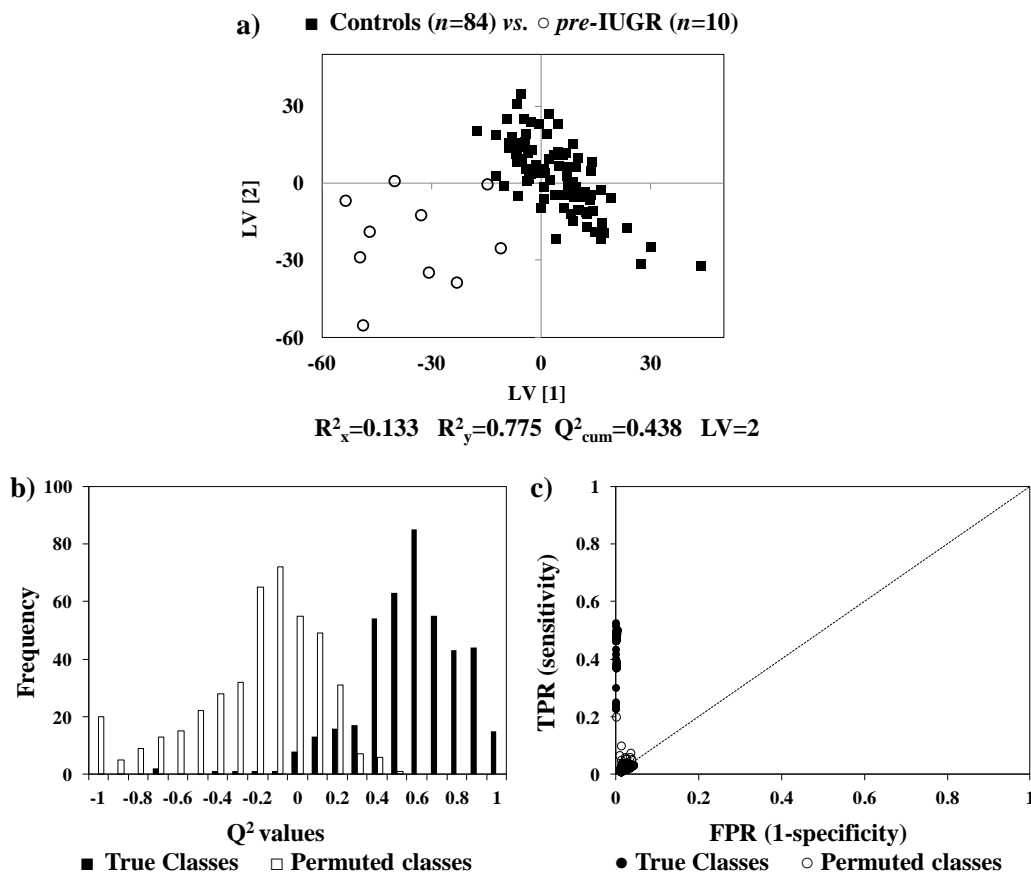


Figure 4.20: a) PLS-DA scores plot, b) Q^2 distributions plot and c) ROC plot of true and permuted classes, given by MCCV obtained for *pre*-IUGR (○, $n=10$) vs. Controls (■, $n=84$).

The variables supporting the separation seen in Figure 4.20, revealed a potentially predictive urinary signature of IUGR. These variables/metabolites are listed in Table 4.7 along with corresponding effect size and p -value. The *pre*-IUGR profile comprised variations in 25 metabolites, 11 unassigned spin systems (singlet at 0.93 ppm, doublet at 1.08 ppm, singlet at 1.82 ppm, 1.99 ppm, 2.03 ppm, 2.05 ppm, 2.37 ppm, 2.91 ppm, quartet at 2.96 ppm, singlet at 4.40 ppm, doublet at 8.78 ppm, broad resonance at 6.92 ppm) and 5

Table 4.7: Metabolite/resonance changes in maternal urine for *pre*-IUGR compared to controls.

^aChemical shifts selected by variables selection; s: singlet, d: doublet, t: triplet, q: quartet, dd: doublet of doublets, m: multiplet, br: broad. ^bMetabolites related to diet and/or gut microflora.

^cChanges specific of the disorder, relatively to other *pre*-diagnosis disorders here evaluated.

^dChanges common to all *pre*-diagnosis disorders. Ui: unassigned compound i following numbering in Table 4.1, Table 4.2, Table 4.4 and Table 4.6. Abbreviations defined in Table 4.1, Table 4.2, Table 4.4 and Table 4.6. Only *p*-values < 0.05 are shown.

<i>pre</i>-IUGR <i>n</i>=10 vs. Controls <i>n</i>=84		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
1-Me-histidine ^b	7.05 s, 7.78 s	↑(0.43)
2-HIBA	1.36 s	↑(0.77)
2-KG	2.45 t, 3.01 t	↓(-0.25)
3-HIVA	1.27 s	↓(-0.50)
3-Me-histidine ^b	7.15 s	↓(-0.29)
4-DEA ^d	1.11 d	↑(0.40)
4-HPA ^b	6.86 d, 7.17 d	↓(-0.35)
4-OH- hippurate ^b	6.98 d, 7.76 d	↓(-0.42)
Acetone ^c	2.23 s	↓(-0.05)
Alanine ^c	1.48 d	↓(-0.38)
Choline ^b	3.20 s	↓(-0.20)
Creatinine ^d	3.05 s, 4.06 s	↓(-0.37)
DMA ^b	2.72 s	↓(-0.43)
Furoylglycine ^b	6.65 q	↓(0.22)
Galactose	5.28 d	↓(-0.43)
Glutamine ^c	2.15 m, 2.47 m	↓(-0.63)
Glycine	3.57 s	↓(-0.57)
Leucine ^c	0.97 t	↓(-0.47)
Lysine ^c	1.73 m, 1.91 m	↓(-0.49)
Succinate	2.41 s	↓(-0.38)
Tartrate ^b	4.34 s	↑(0.93, 4.1×10 ⁻²)
Threonine ^c	1.34 d, 4.26 dd	↓(-0.26)
TMA ^b	2.88 s	↓(-0.36)
Trigonelline ^b	4.44 s, 8.84 br, 9.12 s	↑(0.75)
Tyrosine	6.91 d	↓(-0.48)
U14	0.93 s	↓(-0.30)
U2	1.08 d [2.50, 3.55, 3.72]	↓(-0.13)
U29 ^c	1.82 s	↑(0.41)
U4	1.99 s	↓(-0.56)
U21 ^c	2.03 s	↓(-0.42)
U5 ^c	2.05 s	↓(-0.23)
U34 ^c	2.37 s	↓(-0.35)
U9 ^c	2.91 s	↓(-0.52)
U28 ^c	2.96 q [2.88,4.02]	↓(-0.44)
U10 ^b	4.40 s, 8.78 d	↓(-0.02)
U11 ^c	6.92 br	↓(-0.51)
Unassigned spectral regions		
3.17-3.19		↓(-0.64, 1.9×10 ⁻²)
3.28-3.29		↓(-0.65, 1.6×10 ⁻²)
4.32-4.37		↑(1.01, 4.6×10 ⁻²)
7.00-7.01		↑(1.52, 1.2×10 ⁻²)
7.47-7.49		↑(1.02, 9.2×10 ⁻³)
+ 22 selected regions, <i>p</i> >0.05, at 0-4.6, 6-10 ppm		

unassigned regions with $p < 0.05$. As for previous disorders, these variations were analyzed in terms of specificity, relatively to other *pre*-diagnosed conditions and metabolites found to be specific of *pre*-IUGR are indicated with a ^c while those common to all *pre*-diagnosed disorders are indicated with a ^d in Table 4.7. The VIP-wheel representation of the selected variables enabled the identification of some distinctive features of the *pre*-IUGR profile, as shown in Figure 4.21.

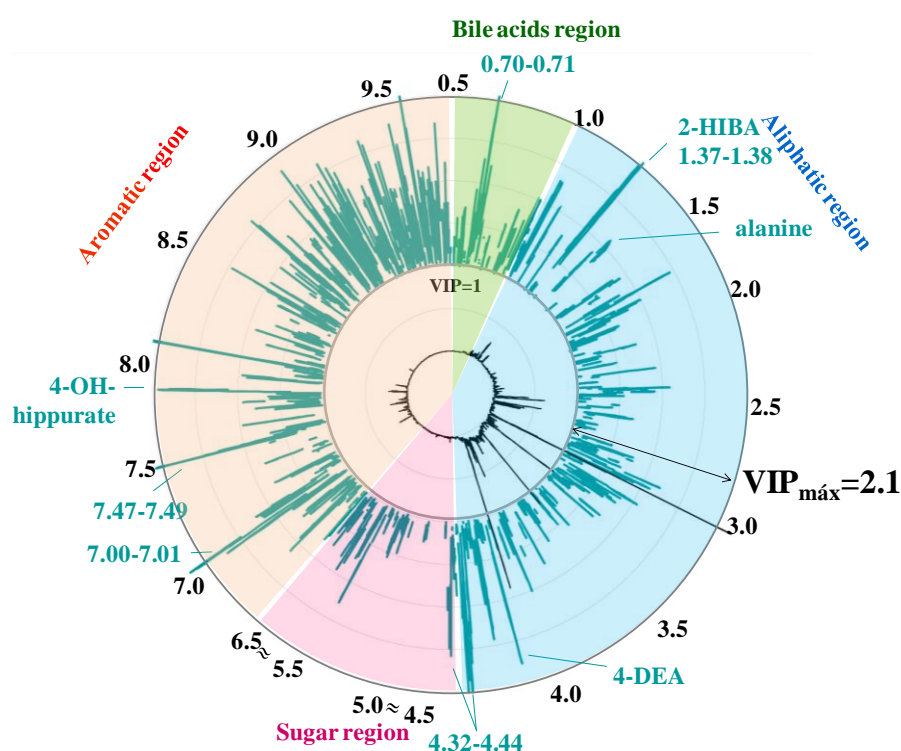


Figure 4.21: VIP-wheel representations of the NMR metabolite signatures obtained for *pre*-IUGR. The average ¹H NMR spectrum of controls is represented in the inner circle, with the corresponding ppm scale shown in the outer black circle. Each dot represents a variable (spectral data point) selected by the intersection method, positioned in a radial VIP scale according to its VIP value. Some distinguishing features are indicated

4.2.3.2 Proposed metabolic interpretation of *pre*-diagnosis urinary changes in IUGR

In this cohort, 7 amino acids were found to be decreased in maternal urine, namely alanine, glutamine, leucine, lysine, threonine, glycine and tyrosine, the first 5 being specific of *pre*-IUGR (Figure 4.22). IUGR is associated with decreased circulating amino

AFFECTED METABOLIC PATHWAYS PRECEDING PTD, PE AND IUGR

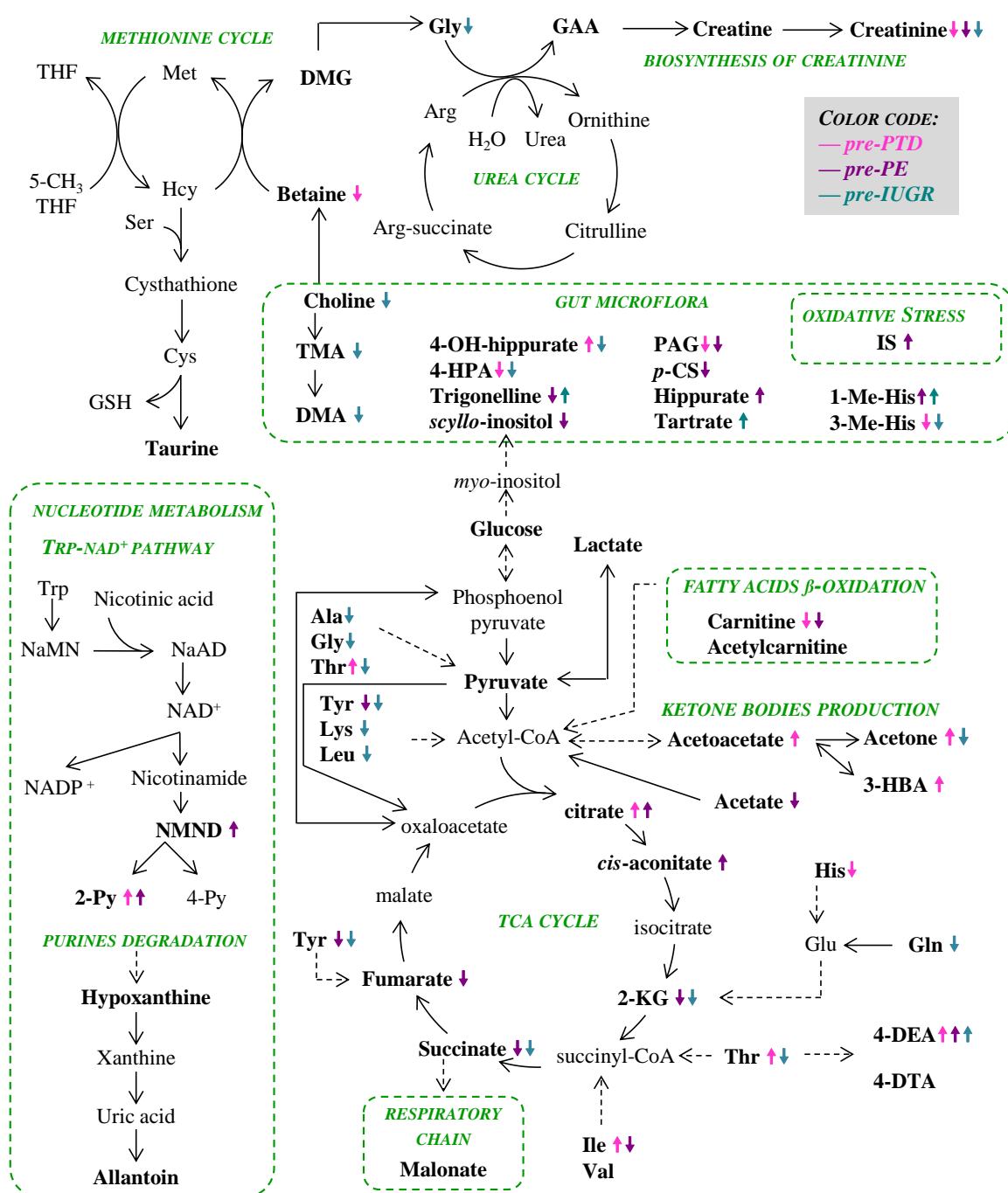


Figure 4.22: Schematic representation of the metabolic pathways possibly affected in *pre*-PTD, *pre*-PE and *pre*-IUGR cases. Metabolites in bold were identified in ¹H NMR spectra of maternal urine. Colored arrows indicate direction of variation in each group. Dashed arrows indicate transformations occurring through several reactions. Amino acids in 3-letter code; PAG: phenylacetylglutamine, *p*-CS: *p*-cresol sulphate, other abbreviations defined in Figure 4.10.

acids and decreased transfer to the fetus through the placenta (Regnault *et al.*, 2005, Horgan *et al.*, 2010, Horgan *et al.*, 2011). The results found in this preliminary study highlight the possibility of detecting a suboptimal amino acid maternal status several weeks before the clinical manifestation of the disease. An interesting question would be to determine if it is the maternal amino acid deficiency that causes the development of the disease, although at this stage, no conclusions can be advanced. Decreased TCA cycle intermediates 2-KG and succinate were also found in *pre*-IUGR samples as well as several metabolites originating from the gut microflora (indicated with a ^b in Table 4.7) although no corroborating findings has been reported in literature. On the other hand, previous evidence of altered lipids metabolism in IUGR after diagnosis (Dessi *et al.*, 2011, Horgan *et al.*, 2011) were not confirmed in this cohort, in the *pre*-diagnosis stage. One final note with regard to the 11 unassigned spin systems, 7 of these being specific of this condition, and 5 spectral regions (with $p < 0.05$) found in the metabolic signature of *pre*-IUGR. It would be of great importance to advance the identification of these compounds so that a more comprehensive metabolic interpretation can be proposed.

4.2.4 Women later developing premature rupture of membranes (PROM)

The final *pre*-diagnosed condition to be studied in this project was premature rupture of the membranes after the 37th g.w. (here named *pre*-PROM). Within the conditions here studied PROM (after the 37th g.w.) poses the lesser impact on maternal and fetal outcomes as this is managed through labor induction without posing significant risks. In this work, maternal urine was collected in the 2nd trimester, 18 to 24 weeks before PROM and compared to controls. A more complete description of samples numbers, metadata is given in Experimental Section 2.1.2. After the initial PLS-DA and MCCV of that model, the ¹H NMR dataset was reduced by variable selection and model was recalculated and resubmitted to validation. The resulting PLS-DA score plot and MCCV parameters are summarized in Table A-III.2, Annex III and shown in Figure 4.23. In this figure a separation trend is visible between controls and *pre*-PROM although with some overlapping of samples.

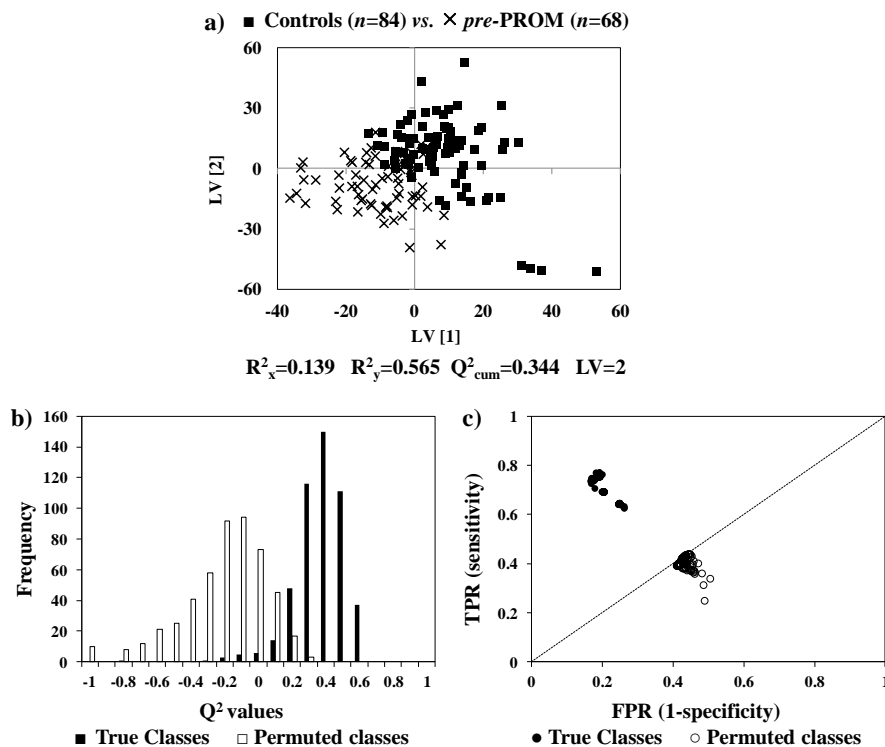


Figure 4.23: a) PLS-DA scores plot, b) Q^2 distributions plot and c) ROC plot of true and permuted classes, given by MCCV obtained for *pre*-PROM (\times , $n=68$) vs. Controls (\blacksquare , $n=84$).

Regarding the MCCV results, in spite of the low overlapping of Q^2 of true and permuted models, confusion matrices yielded a CR of 70%, sensitivity of 68% and specificity of 72%, these values being significantly lower than those obtained for the remaining disorders. Thus, this group was found to be indistinguishable of controls in terms of their 2nd trimester urinary metabolome.

4.3 Summary of the impact of fetal and maternal disorders

A summary of the main biochemical findings of fetal and maternal disorders here studied on 2nd trimester maternal urine composition is shown in Figure 4.24, along with the corresponding validation parameters in decreasing order of impact. As shown in Figure 4.24, it was found that T21 compared to other CD types caused the highest impact, with CR, sensitivity and specificity ranging from 92-95%, followed by the CNS compared to other FM types, which had CR, sensitivity and specificity of 83-90%. These results support the premise of a possible non-invasive diagnostic procedure of T21 and CNS malformations based on maternal urine metabolomics. The impact of general CD and FM

cases studied here through the consideration of heterogeneous groups, were associated with slightly lower CR, sensitivity and specificity (85%, 87%, 84% and 88%, 74%, 87%, respectively for CD and FM). These lower validation parameters obtained for these disorders may be a consequence of the heterogeneous nature of the groups used (comprising several CD and FM types). Nevertheless, in spite of the relatively lower CR, sensitivity and specificity, the results here obtained are still higher than 2nd trimester non-invasive diagnostic methods presently available in clinical practice (CR 17-74% for FM by ultrasound, depending on the population and operator skill, and 60-75% for T21-no reference for heterogeneous CDs were found).

Disorder studied	Main biochemical findings	Validation parameters
T21	Methionine and nucleotide metabolism, increased oxidative stress and perturbations in the respiratory chain	CR: 94% Sensitivity: 92% Specificity: 95%
CNS	Methionine and amino acids metabolism, increased oxidative stress, perturbations in the respiratory chain and enhanced TCA cycle	CR: 87% Sensitivity: 83% Specificity: 90%
CD	Enhanced ketone bodies production, derangements in sugars, amino acids and nucleotide metabolism, enhanced TCA cycle operation, urea cycle and/or creatine biosynthesis	CR: 85% Sensitivity: 87% Specificity: 84%
FM	Enhanced gluconeogenesis and TCA cycle, perturbation in the respiratory chain, increased lipids oxidation and ketone bodies production and perturbations in nucleotide metabolism	CR: 88% Sensitivity: 74% Specificity: 87%
pre-PTD	Perturbations in amino acid metabolism and/or placental transfer and increased lipids oxidation and oxidative stress	CR: 84% Sensitivity: 79% Specificity: 84%
pre-IUGR	Decreased amino acids levels	CR: 94% Sensitivity: 99% Specificity: 94%
pre-PE	Perturbations in TCA cycle	CR: 94% Sensitivity: 75% Specificity: 94%
pre-PROM	—	CR: 70% Sensitivity: 68% Specificity: 72%

Figure 4.24: Summary of the main biochemical findings of fetal and maternal disorders and corresponding validation parameters in decreasing order of impact.

Regarding the prediction of later developed disorders the highest impact was found for PTD, followed by IUGR, PE and PROM. The *pre*-PTD model was associated with CR 84%, sensitivity 79% and specificity 84%. The low sensitivity of this model expressing a low true positive rate (TPR), meaning that some women that later develop PTD are misclassified as controls. A possible explanation for this limitation may be the interval between sampling and diagnosis (11-20 g.w.). The *pre*-IUGR and *pre*-PE models, in spite of their exploratory nature due to very low number of samples available ($n < 10$), yielded good validation results (all above 94% with exception of the sensitivity of the *pre*-PE model (75%), i.e. low TPR). The low sensitivity or TPR of *pre*-PE may be a direct consequence of the low number of samples used and/or the interval from sampling to diagnosis (14-21 g.w.). Lastly, the *pre*-PROM condition was here associated with CR, sensitivity and specificity ranging from 68-72%, thus considered indistinguishable of controls. The absence of changes in *pre*-PROM may be due to the lower impact of this pregnancy disorder and/or the interval from sampling to diagnosis (18-24 g.w.).

5. Gestational Diabetes Mellitus: a deeper insight into maternal metabolism disturbances

This chapter describes a study of the metabolic effects of gestational diabetes mellitus (GDM), at several disease stages, through urine metabolomics. Hence, maternal urine NMR profiling was carried out at three different stages: a) *Diagnosed GDM* (samples collected at the time of diagnosis, prior to treatment initiation), b) *pre-diagnosis* (samples collected at 2nd trimester, 2-21 weeks prior to the clinical manifestation of the disease) and c) *Treated GDM* (samples collected during GDM treatment). Firstly, the urinary metabolic signature of each GDM stage was unveiled through comparison with healthy gestational age (GA) matched controls. Additionally, the study of the metabolic effect of insulin and non-insulin treatments was sought, by comparing *insulin-treated* and *non-insulin-treated* subjects. Secondly, a novel pharmacometabolomic application for GDM treatment was attempted, by exploring the possibility to predict future insulin treatment requirement. These results demonstrate, for the first time, the enticing possibility of unveiling a metabolic phenotype potentially predictive of individual future GDM treatment requirements.

5.1 Maternal urine composition at the time of GDM diagnosis

5.1.1 Unveiling an urinary metabolic signature of GDM

The first disease stage to be characterized in the GDM study was the diagnosed state, i.e. samples collected at the time of diagnosis, at 16-33 g.w., prior to treatment initiation, in order to broaden the understanding of the metabolic effect of the disease in its more exacerbated state. Detailed description of sample metadata and diagnostic criteria is shown in Section 2.1.2 and Table 2.4 of Experimental Section.

The average ¹H NMR spectra of maternal urine collected at each disease stage are shown in Figure 5.1, along with some relevant assignments. Visually identifiable differences in the average composition of urine, are indicated in Figure 5.1c, namely in the aliphatic region from 1.0-1.5 ppm and 1.9-2.5 ppm as well as the aromatic region from 7-8 ppm.

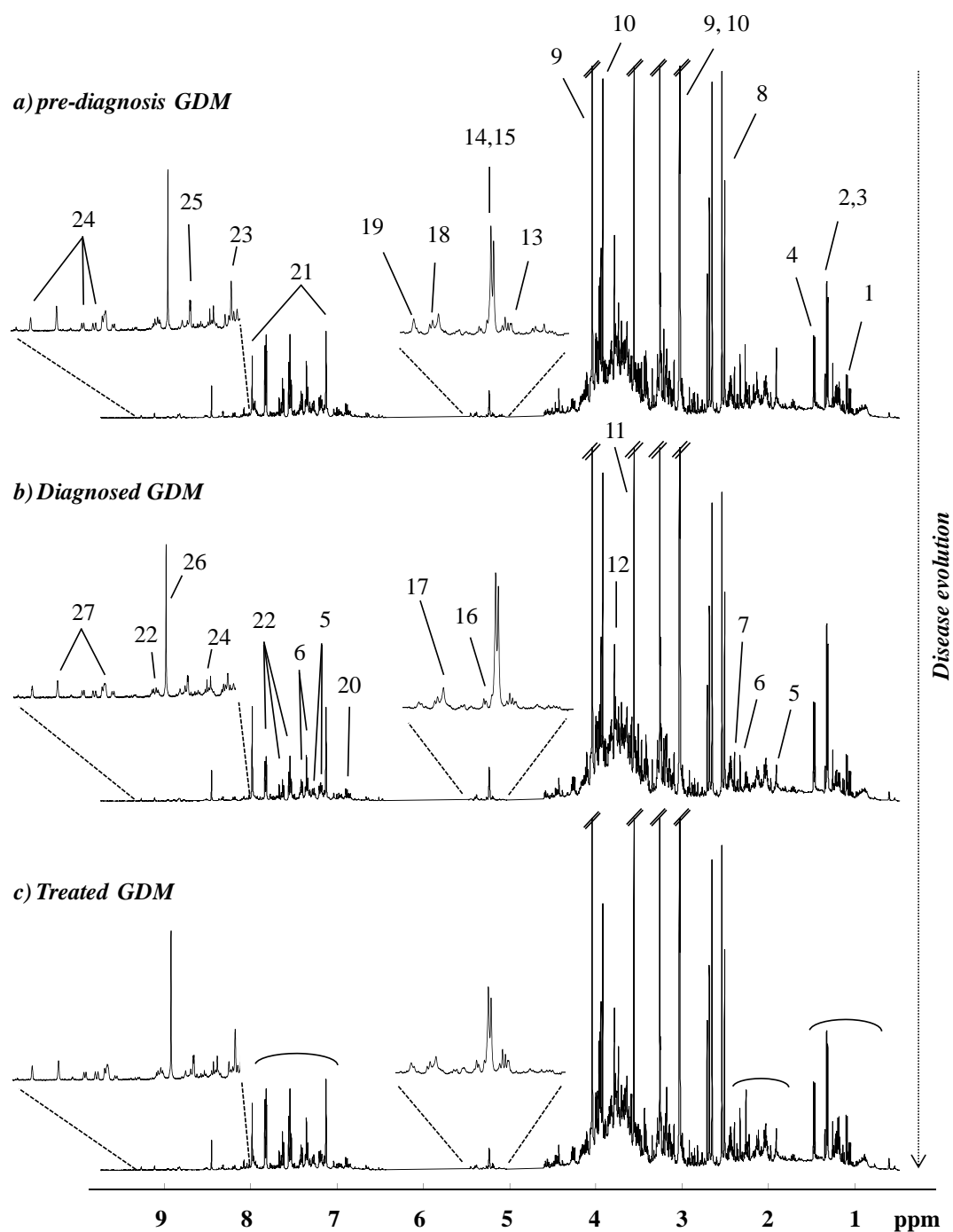


Figure 5.1: Average 500 MHz ^1H NMR spectra of a) *pre-diagnosis* (median GA 17 g.w.), b) *Diagnosed* (median GA 27 g.w.) and c) *Treated* GDM (median GA 35 g.w.) groups. Some relevant assignments are noted in a,b) and curved lines in c) indicate regions with major visual changes. Legend (3-letter code for a.a.): 1: 4-deoxyerythronic acid (4-DEA), 2: thr, 3: lactate, 4: alanine, 5: acetate, 6: phenylacetylglutamine (PAG), 7: *p*-cresol sulphate (*p*-CS), 8: citrate, 9: creatinine, 10: creatine, 11: gly, 12: guanidoacetate (GAA), 13: xylose, 14: glucose, 15: lactose, 16: galactose, 17: allantoin, 18: sucrose, 19: 1,6-anhydroglucose, 20: tyr, 21: his, 22: hippurate, 23: 3-methylhistidine, 24: hypoxanthine, 25: *N*-methyl-2-pyridone-5-carboxamide (2-Py), 26: trigonelline, 27: *N*-methylnicotinamide (NMND). Vertical arrow indicates disease evolution from *pre-diagnosis*, *diagnosed* and *treated* stages.

In order to unveil the urinary signature of GDM, samples collected at the time of diagnosis were compared with controls. The *Diagnosed GDM* group comprises samples collected during the 2nd ($n=8$) and 3rd ($n=5$) trimesters of pregnancy (GA range from 16-33 g.w.). As gestational age was found to cause a significant impact on the composition of 2nd and 3rd trimester maternal urine, these differences having been discussed Chapter 3, a control group matched, as much as possible, in number (approximately the same number of samples) and GA (2nd T, $n=6$, 3rd T, $n=8$, GA range of 16-33 g.w.) was selected, here named *control group 1*. For this study, a total of 113 2nd and 25 3rd trimester controls were available. However, if all control samples (2nd and 3rd trimester) were used, the resulting control group would be significantly unbalanced in terms of GA and the resulting urinary profiles would reflect mainly the GA rather than GDM. So, in order to avoid potentially erroneous results the control group was selected based solely on number and GA at sampling. The resulting score plot of the PLS-DA model, obtained with the full resolution spectral data, and corresponding MCCV results is shown in Figure 5.2 (left). In spite of the separation trend between GDM and controls in the score plot (Figure 5.2a), there is significant dispersion within the two groups. Moreover, MCCV yielded marginally satisfactory results with CR 77%, sensitivity 74% and specificity 81%, and some overlap of Q^2 distribution of true and permuted models (Figure 5.2c,e and Annex III, Table A-III.3). Thus, the ^1H NMR dataset was reduced (using the variable selection method presented in Chapter 4), PLS-DA was recalculated using the resulting subset of variables and these were resubmitted to MCCV (Figure 5.2 right). A significant reduction in sample dispersion became visible in the new PLS-DA score plot shown Figure 5.2b, this model having very good prediction power, supported by the almost absent overlap of Q^2 distributions, as well as high robustness, with CR 96%, sensitivity 93% and specificity 100% (Figure 5.2d,f, Table A-III.3).

In the PLS-DA score plot shown in Figure 5.2b, two circles indicate one control and one *Diagnosed GDM* sample both collected at 16 g.w., i.e. the minimum GA in both groups. The location of this control sample in negative LV2 relatively to other controls, whilst the GDM sample is clustered with the rest of the disorder group, suggests that the GA effect is significant in healthy pregnancies (consistent with findings presented in Chapter 3.3), but not so much for the GDM sample. In fact, the position of the latter

(collected at the 16th week of gestation) seems to be determined by a particular profile specific of the disease and not of GA.

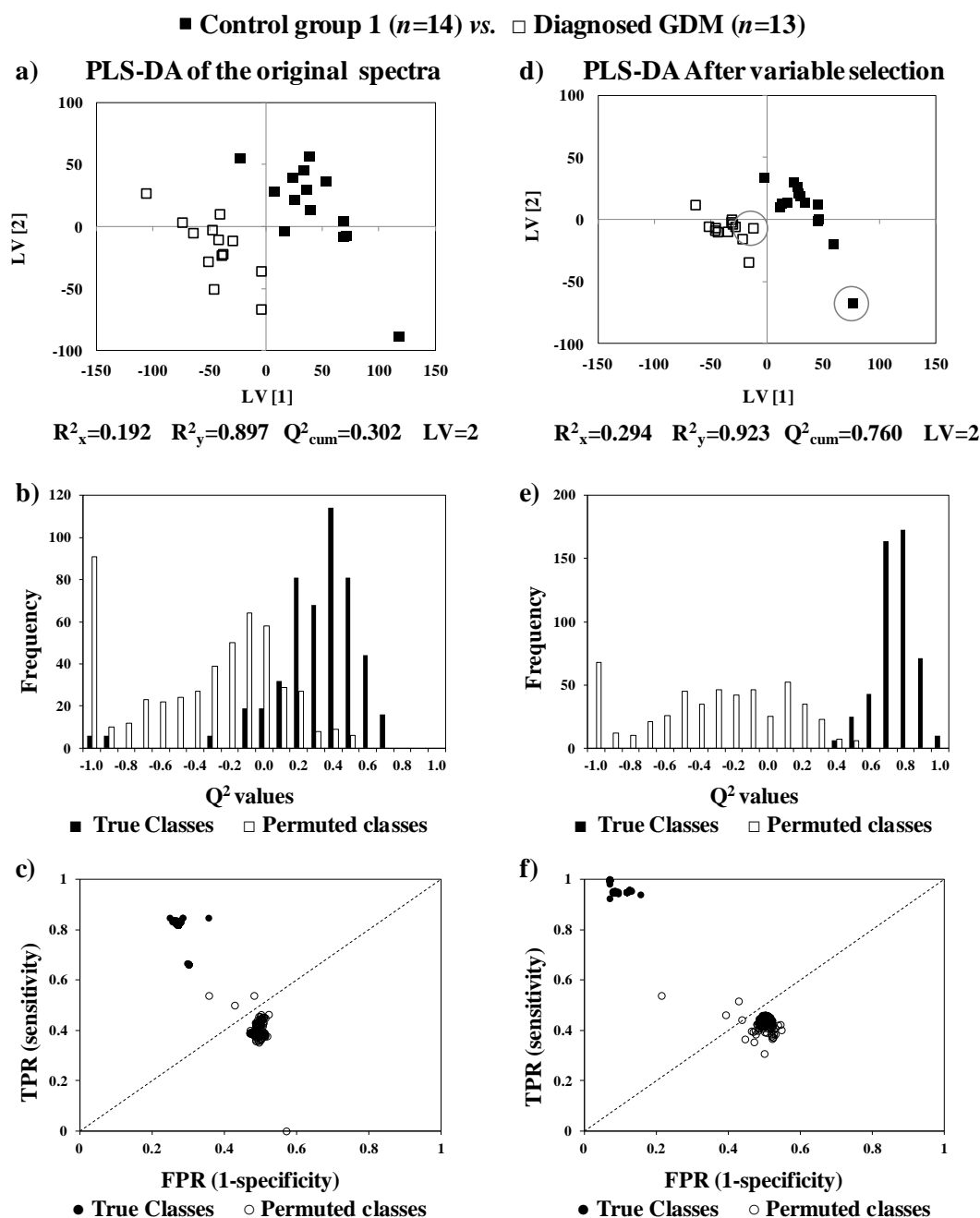


Figure 5.2: PLS-DA score plots (a,d), Q^2 (b,e) and ROC plots (c,f) of true and permuted models obtained for the original ^1H NMR spectra of urine of *Diagnosed GDM* ($n=13$) vs. *control group 1* ($n=14$) (left) and for the same spectra after variable selection method presented in section 4.1.1.1 (right). TPR: true positive rate, FPR: false positive rate. Circle indicates two samples collected at 16 g.w. (minimum GA of both groups) showing clustering of the GDM group due to disease effect and higher distance in the control group due to GA, consistent with findings previously shown in Chapter 3.3.

The analysis of the subset of variables explaining the separation obtained through PLS-DA enabled a metabolic signature of GDM to be identified and characterized in terms of effect size and *p*-value (Table 5.1). The resulting signature profile comprised variations in 18 metabolites (including sugars, organic acids and amino acids), 4 unassigned spin systems (doublet at 1.25 ppm and singlets at 2.02 ppm, 2.03 ppm and 7.06 ppm) and 10

Table 5.1: Metabolite/resonance changes in maternal urine of women diagnosed with GDM compared to controls. ^a Chemical shifts selected by variables selection; TOCSY or STOCYSY correlated peaks in square brackets; s: singlet, d: doublet, t: triplet, q: quartet, dd: doublet of doublets, m: multiplet, br: broad. ^b Peaks possibly related to diet and/or gut microflora. ^c Consistent with previous findings in maternal urine (Sachse *et al.*, 2012). Ui: unassigned compound i by appearance order. 2-KG: 2-ketoglutarate, 4-HPA: 4-hydroxyphenylacetate, DMA: dimethylamine, DMG: dimethylglycine, *p*-CS: *p*-cresol sulphate. Only *p*-values<0.05 are shown.

Diagnosed GDM (n=13) vs. Control group 1 (n=14)		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
1,6-anhydroglucose ^b	5.46 br	↓(-0.65)
2-KG	2.45 t	↑(1.22, 7.80×10 ⁻³)
4-HPA ^b	7.17 d	↓(-0.27)
4-OH-hippurate ^b	6.98 d, 7.76 d	↓(-0.26)
Betaine	3.26 s	↑(0.84, 2.55×10 ⁻²)
Citrate ^c	2.54 d, 2.69 d	↑(0.47)
Creatine ^b	3.04 s, 3.94 s	↑(1.41, 3.95×10 ⁻³)
Creatinine ^b	3.05 s, 4.06 s	↓(-0.25)
DMA ^b	2.72 s	↓(-0.35)
DMG	2.93 s	↑(1.00, 1.45×10 ⁻²)
Furoylglycine ^b	6.65 q	↓(-0.42)
Glycine	3.57 s	↑(0.19)
Glutamine	2.47 m	↑(1.01)
Hippurate ^b	3.97 d, 7.56 t, 7.64 t, 7.83 d, 8.52 br	↓(-1.20, 1.68×10 ⁻²)
<i>p</i> -CS ^b	2.35 s, 7.29 d	↓(-0.57)
Sucrose ^b	5.41 d	↓(-0.77)
Succinate	2.41 s	↑(0.94, 2.91×10 ⁻²)
Threonine	4.26 dd	↑(0.98, 2.22×10 ⁻²)
U1	1.25 d [3.88]	↓(-0.99, 4.27×10 ⁻²)
U2	2.02 s	↑(0.89, 5.43×10 ⁻²)
U3	2.03 s	↑(0.70, 9.45×10 ⁻²)
U4	7.06 s	↓(-1.17, 5.60×10 ⁻³)
Unassigned spectral regions		
2.91-2.93		↑(1.24, 6.61×10 ⁻⁴)
2.95-2.98		↑(1.10, 1.25×10 ⁻²)
3.18-3.23		↑(0.95, 2.22×10 ⁻²)
3.81-3.82		↑(0.85, 4.27×10 ⁻²)
5.11-5.13		↓(-1.55, 3.72×10 ⁻⁵)
5.36-5.37		↓(-1.00, 2.91×10 ⁻²)
5.47-5.49		↓(-1.27, 6.62×10 ⁻³)
7.47-7.50		↓(-1.10, 6.62×10 ⁻³)
8.14-8.17		↓(-2.02, 9.67×10 ⁻⁶)
8.23-8.32		↓(-1.69, 5.28×10 ⁻⁴)
+ 15 unassigned spin systems and 3 spectral regions with <i>p</i> >0.05		

spectral regions (i.e. regions comprising unassignable/overlapped resonances) with individual significance ($p < 0.05$). Moreover, metabolite variations found consistent with previous reports are indicated with ^c (found only for citrate) while those arising from gut microflora and/or diet are noted with ^b in Table 5.1.

5.1.2 Proposed metabolic interpretation of urinary changes in GDM

A schematic representation of the main metabolic pathways affected in the *Diagnosed GDM* group is shown in Figure 5.7 (page 199), along with changes in other GDM states discussed throughout this chapter. Amino acids glycine, glutamine and threonine were found increased in GDM, the latter with individual significance. Increased glycine has been found in plasma of GDM-affected women (Pappa *et al.*, 2007) and increased glutamine in urine of T2DM (Salek *et al.*, 2007) both variations being related to enhanced gluconeogenesis, this enhancement also explaining the concomitant increase in threonine found in this study. TCA cycle intermediates citrate, 2-ketoglutarate (2-KG) and succinate were also found to increase, the last two with individual significance. Citrate has been found to increase in urine of women diagnosed with GDM (Sachse *et al.*, 2012), but also in urine of T2DM (Salek *et al.*, 2007) and T1DM (Deja *et al.*, 2013), whilst succinate and 2-KG have been reported to decrease (opposite direction) in urine of T2DM (Salek *et al.*, 2007). Moreover, increased urinary excretion of TCA intermediates has been associated with altered kidney function or systemic stress produced by hyperglycemia in T2DM (Salek *et al.*, 2007). Thus, the variations found in this cohort are consistent with enhanced TCA cycle operation, along with a possible contribution of altered renal function, as occurs in T1DM and T2DM. Betaine and dimethylglycine (DMG), both metabolites involved in methionine metabolism, were found significantly increased in GDM women. Betaine is metabolized from choline, and donates a methyl group to homocysteine to form methionine, also generating DMG which is then converted to glycine (Dasarathy *et al.*, 2010). Changes in methionine metabolism has been associated with several pregnancy related disorders, namely with the occurrence of fetal malformations, preeclampsia, spontaneous abortion, placental abruption and premature delivery (Dasarathy *et al.*, 2010), but not to GDM. Moreover, increased betaine and DMG have been linked with choline degradation in the gut microflora or due to altered

lipoprotein turnover/biosynthesis in T2DM (Salek *et al.*, 2007). Decreased creatinine and increased creatine were also found in this cohort, only the latter being individually significant. Creatinine is formed from creatine, which is obtained either by diet or from conversion of guanidoacetate (GAA) during the urea cycle (Brosnan and Brosnan, 2010). Moreover, creatine and creatinine have been reported to decrease in urine of T2DM due to changes in the glomerular filtration rate (Salek *et al.*, 2007). The variations found in this cohort suggest a perturbation in creatinine synthesis, supported by the unbalanced creatine/creatinine excretion, along with a possible contribution of renal function alterations.

Metabolites deriving from gut microflora and/or diet were also found to vary in this study, particularly hippurate, which is significantly decreased in GDM women, along with decreased 4-hydroxyphenylacetate (4-HPA), 4-OH-hippurate, *p*-cresol sulphate (*p*-CS), dimethylamine (DMA), furoylglycine and sugars 1,6-anhydroglucose and sucrose. Out of these metabolites, only hippurate has been related with diabetes, although different reports differ regarding the direction of variation of this metabolite. Two studies have reported decreased urinary hippurate in relation to T2DM (Salek *et al.*, 2007) and T1DM (Deja *et al.*, 2013), whilst one other study reported increased hippurate in urine of T1DM (Zuppi *et al.*, 2002). Regarding the expected glucosuria in GDM cases, a consequence of hyperglycemia as in other types of diabetes, such was not found in this work. Visual inspection of the NMR spectra revealed that only 2 out of the 13 GDM samples contained high glucose levels, thus not representing the bulk of the group. For this reason, urinary glucose was not found to be part of the metabolic signature of the GDM, at this stage, although larger cohorts are needed to confirm this finding. Moreover, it is known that urinary excretion of glucose in pregnancy occurs due to adaptations in renal function (Hyttén, 1973, Creasy *et al.*, 2009) not necessarily indicating a pathological situation.

It is important to note again that all the above changes should be confirmed in a larger cohort ($n \gg 13$), although the results presented here are a good starting point for the recognition of exacerbated GDM effects.

5.2 Women later developing GDM

5.2.1 Unveiling a *pre-diagnosis* urinary metabolic signature of GDM

For this study, urine of women diagnosed GDM 2-21 g.w. after sampling, prior to any clinical manifestation of the disease (named *pre-diagnosis GDM*) was compared to a GA-matched controls (*control group 2*). As GA at sampling for all 2nd trimester controls (15-26 g.w.) was approximately the same as that of the *pre-diagnosis* GDM group (15-22 g.w.), all samples were used for the comparison. A more detailed description on sample numbers and metadata is shown in Section 2.1.2 and Table 2.4 of Experimental Section.

The comparison of the original datasets of *pre-diagnosis GDM* samples and controls through PLS-DA revealed a separation trend in the score plot (not shown) although MCCV showed that the model was neither robust nor had predictive power (see model and MCCV parameters in Annex III, Table A-III.3.). Thus, the ¹H NMR dataset was compressed through variable selection and PLS-DA recalculated and tested through MCCV. The resulting model revealed a separation trend between controls and disease in the score plot, as shown in Figure 5.3a, as well as good predictive power, given by the Q^2 distribution of true and permuted models, and robustness, with CR 87%, sensitivity 84% and specificity 89%, as shown in Figure 5.3b,c and Table A-III.3.

The analysis of the subset of variables explaining the separation seen though PLS-DA enabled a potentially predictive metabolic signature to be unveiled (Table 5.2). This signature comprises variations in 24 metabolites (including amino and organic acids, ketone bodies, sugars and metabolites derived from the gut microflora), 5 unassigned spin systems (doublet at 1.08 and 1.15 ppm, singlets at 1.99 and 1.82 ppm, and a triplet at 4.30 ppm) and 15 spectral regions with $p < 0.05$. Metabolite variations here found consistent with previous findings (including two papers published as a result of this work) are indicated with ^c in Table 5.2, namely higher 3-hydroxyisovalerate (3-HIVA), 4-deoxyerythronic acid (4-DEA), acetate, choline, citrate, glucose, *N*-methylnicotinamide (NMND), trimethylamine (TMA), xylose, unassigned doublet at 1.08 ppm and lesser 4-HPA, creatinine, DMA, hippurate, *p*-CS, phenylacetylglutamine (PAG), unassigned doublet at 1.15 ppm and singlets at 1.99 and 1.82 ppm. In the same table, changes common to other

pre-diagnosis disorders, described in Chapter 4.2, are noted with ^d (higher 4-DEA and lesser creatinine) whilst ^b denotes metabolites arising from diet and/or gut microflora.

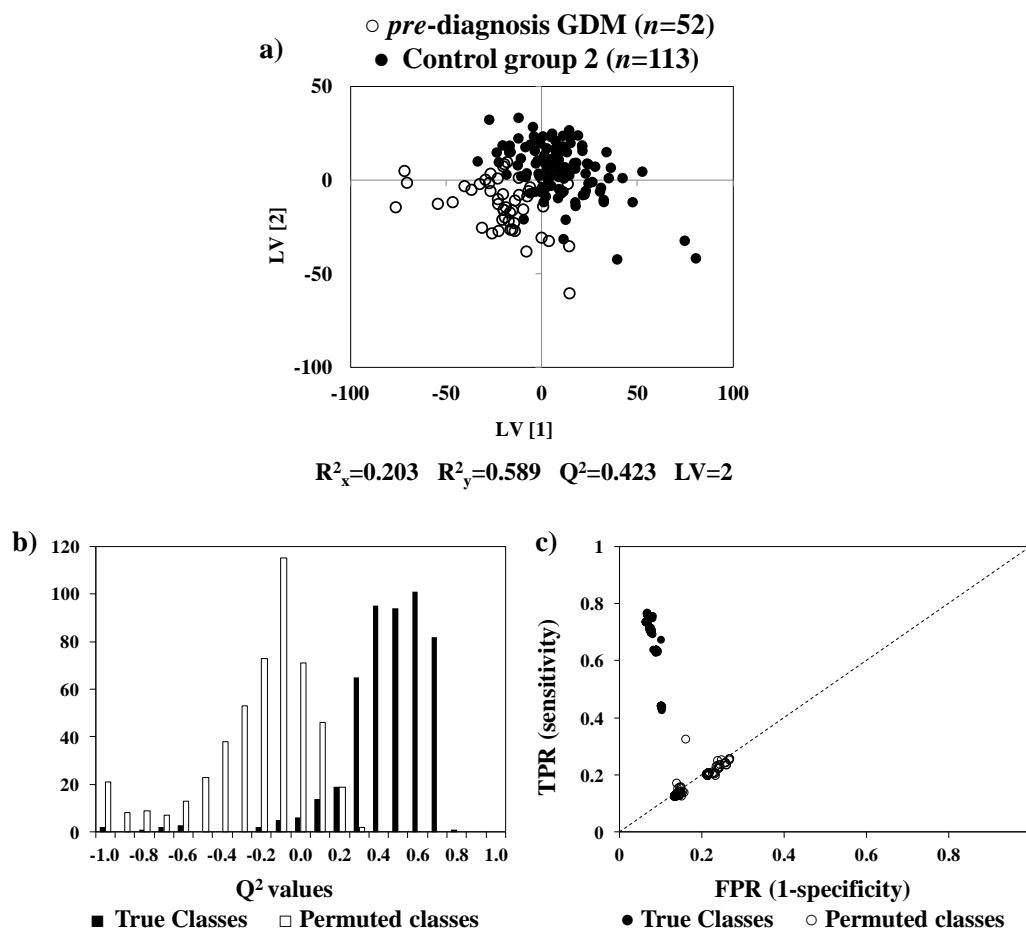


Figure 5.3: a) PLS-DA score plots, b) Q^2 and c) ROC plots of true and permuted models obtained for *pre*-diagnosis GDM (n=52) vs. control group 2 (n=113). TPR: true positive rate, FPR: false positive rate.

The comparison of the urinary profile of *Diagnosed* and *pre*-diagnosed GDM groups revealed that 7 changes are common between these two groups (citrate, threonine, hippurate, *p*-CS, 4-HPA, DMA and creatinine). In addition, 4-OH-hippurate was found increased in *pre*-diagnosis and decreased at diagnosis. The remaining 16 metabolites seem to be specific of the *pre*-diagnosed state, these comprising 4 amino acids and derivatives, 3 lipid oxidation metabolites, 2 sugars, 2 nucleotide metabolism products, 3 gut microflora metabolites and 2 others (3-HIVA and acetate). The above changes support the value of the methodology to predict GDM development, allowing earlier clinical intervention.

Table 5.2: Metabolite/resonance changes in maternal urine for *pre-diagnosis GDM* compared to *control group 2*. ^aChemical shifts selected by variables selection; TOCSY or STOCYSY correlated peaks in square brackets; s: singlet, d: doublet, t: triplet, dd: doublet of doublets, br: broad. ^bPeaks possibly related to diet and/or gut microflora. ^cConsistent with previous findings in maternal urine (Diaz and Pinto *et al.*, 2011, Graça *et al.*, 2012, Sachse *et al.*, 2012, Diaz *et al.*, 2013). ^dChanges common to all *pre*-diagnosed disorders presented in Chapter 4.2. Ui: unassigned compound i ordered following Table 5.1. 3-HBA: 3-hydroxybutyrate, 3-HIVA: 3-hydroxyisovalerate, 4-DEA: 4-deoxyerythronic acid, 4-DTA: 4-deoxythreonic acid, GAA: guanidoacetate, NMND: *N*-methylnicotinamide, PAG: phenylacetylglutamine, TMA: trimethylamine, other abbreviations defined in Table 5.1. Only *p*-values<0.05 are shown.

<i>pre</i>-diagnosis GDM (<i>n</i>=52) vs. Control group 2 (<i>n</i>=113)		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
3-HBA	1.20 d	↑(0.39, 3.31×10 ⁻²)
3-HIVA ^c	1.27 s, 2.37 s	↑(0.31, 9.75×10 ⁻²)
4-DEA ^{c, d}	1.11 d	↑(0.56, 2.82×10 ⁻³)
4-DTA	1.23 d	↑(0.20, 3.97×10 ⁻²)
4-HPA ^{b, c}	6.86 d, 7.17 d	↓(-0.41, 3.13×10 ⁻³)
4-OH-hippurate ^b	3.95 s, 6.98 d	↑(0.31)
Acetate ^c	1.93 s	↑(0.40)
Acetoacetate	2.29 s, 3.46 s	↑(0.35)
Allantoin	5.39 s	↓(-0.14)
Carnitine ^c	3.23 s	↓(-0.23)
Choline ^{b, c}	3.20 s	↑(0.34, 7.28×10 ⁻²)
Citrate ^c	2.54 d, 2.69 d	↑(0.39, 9.47×10 ⁻²)
Creatinine ^{b, c, d}	4.06 s	↓(-0.18)
DMA ^{b, c}	2.72 s	↓(-0.30)
Glucose ^c	5.25 d	↑(0.74, 1.90×10 ⁻²)
GAA	3.80 s	↑(0.16)
Hippurate ^{b, c}	3.97 d, 7.56 t, 7.64 t, 7.83 d, 8.52 br	↓(-0.39, 6.39×10 ⁻⁴)
Histidine	4.01 dd	↑(0.10)
NMND ^c	4.48 s, 8.90 d, 8.97 d, 9.29 s	↑(0.57, 3.96×10 ⁻³)
<i>p</i> -CS ^b	7.21 d, 7.29 d	↓(-0.34)
PAG ^{b, c}	2.27 t, 7.35 t, 7.43 t	↓(-0.45, 9.99×10 ⁻³)
Threonine	4.26 dd, 3.60 d	↑(0.35, 3.61×10 ⁻²)
TMA ^{b, c}	2.88 s	↑(0.45)
Xylose ^{b, c}	4.59 d, 5.21 d	↑(0.22, 1.62×10 ⁻²)
U5 ^c	1.08 d [2.50, 3.55, 3.72]	↑(0.36, 1.21×10 ⁻²)
U6 ^c	1.15 d [3.45, 3.55, 3.89]	↓(-0.29, 8.60×10 ⁻²)
U7 ^c	1.99 s	↓(-0.31, 6.48×10 ⁻²)
U8 ^c	1.82 s	↓(-0.26, 4.77×10 ⁻²)
U9	4.30 t [4.18]	↑(0.46, 3.23×10 ⁻³)
Unassigned spectral regions		
0.81-0.82		↓(-0.35, 7.22×10 ⁻²)
1.94-1.96		↓(-0.36, 2.34×10 ⁻²)
3.24-3.26		↑(0.66, 1.62×10 ⁻³)
3.46-3.52		↑(0.75, 3.02×10 ⁻³)
3.53-3.56		↑(0.43, 3.46×10 ⁻³)
3.66-3.69		↓(-0.41, 1.22×10 ⁻²)
3.71-3.76		↑(0.48, 1.78×10 ⁻²)
3.77-3.80		↑(0.27, 8.79×10 ⁻²)
3.83-3.86		↑(0.63, 9.21×10 ⁻³)
3.89-3.90		↑(0.39, 7.34×10 ⁻²)
3.92-3.93		↑(0.46, 6.58×10 ⁻²)
4.14-4.16		↑(0.40, 3.79×10 ⁻³)
4.17-4.21		↓(-0.30, 2.15×10 ⁻²)
6.72-6.78		↓(-0.59, 7.12×10 ⁻⁵)
7.08-7.09		↓(-0.36, 1.04×10 ⁻²)
+ 11 unassigned spin systems and 9 spectral regions with <i>p</i> >0.05		

5.2.2 Proposed metabolic interpretation of the *pre*-diagnosis changes in GDM

The main changes found for the *pre*-diagnosed group compared to controls are noted in the schematic representation of the affected metabolic pathways shown in Figure 5.7 (page 199). Regarding amino acid metabolism, a significant increase was found for threonine, along with its catabolic products 4-deoxythreonic acid (4-DTA) and 4-DEA (structures shown in Figure 5.4), and for histidine. The increased excretion of 4-DTA and 4-DEA was related to ketone bodies production in the 3rd trimester of healthy pregnancy, determined through STOCSY (as presented and discussed in Chapter 3.3), possibly due to the ketogenic nature of threonine. Consistently with the changes in ketone bodies production, 3-HBA and acetoacetate levels were found to increase for the first time in this work. Moreover, the tandem increase in ketone bodies and decreased carnitine, also suggest increased β -oxidation of fatty acids, a condition known to occur in T1DM and T2DM (Messana *et al.*, 1998, Zuppi *et al.*, 2002, Salek *et al.*, 2007, Suhre *et al.*, 2010), and found here for the first time to precede GDM diagnosis.

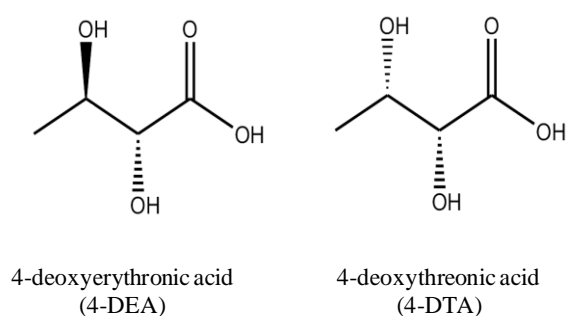


Figure 5.4: Chemical structure of 4-deoxyerythronic acid (4-DEA) and 4-deoxythreonic acid (4-DTA).

Increased citrate in both *Diagnosed* and *pre-diagnosed* states, and consistently with previous reports, one published in the context of this thesis, (Sachse *et al.*, 2012, Diaz *et al.*, 2013), highlights the relevance of this metabolite for GDM urinary signature. Higher citrate excretion had been found in previous *pre*-diagnosis GDM studies, in T2DM and T1DM possibly due to changes in renal function (Salek *et al.*, 2007, Deja *et al.*, 2013). Decreased creatinine was found in the *pre*-diagnosis state, as in the *Diagnosed GDM* group and in previous T2DM reports also due to changes in renal function (Salek *et al.*, 2007). Nonetheless, decreased creatinine was also found in all *pre*-diagnosis disorders studied in

this work (*pre*-PTD, *pre*-IUGR and *pre*-PE, presented and discussed in Chapter 4.2) possibly reflecting common stress effect, thus compromising its individual value for specific disease prediction.

A notable finding refers to metabolites deriving from the gut microflora, namely for choline, TMA, DMA, hippurate, PAG, 4-HPA, 4-OH-hippurate and *p*-CS. This finding emphasizes the involvement of the gut microflora in the development of the disease (although the exact nature remaining unknown) and reinforces the importance of these metabolites in the definition of a potentially predictive profile for GDM. The involvement of the gut microflora in GDM is consistent with previous *pre-diagnosis* GDM reports, published as result of this work (Diaz *et al.*, 2013), *pre*-T2DM (Chen *et al.*, 2008, Zhao *et al.*, 2010), T2DM (Messana *et al.*, 1998, Salek *et al.*, 2007, Zhang *et al.*, 2009, Suhre *et al.*, 2010) and T1DM (Lanza *et al.*, 2010, Culeddu *et al.*, 2012, Deja *et al.*, 2013) reports. Out of the metabolites involved in the gut microflora and found varying in this cohort, hippurate and PAG gain additional relevance as both metabolites decrease with individual significance ($p < 0.05$), and both having been previously found decreased in *pre-diagnosis* GDM (Diaz *et al.*, 2013), and *pre*-T2DM (Zhao *et al.*, 2010). Another important variation preceding the development of GDM is the choline increase, which had been previously found in maternal urine, using smaller cohorts and analyzed by ^1H NMR and UPLC-MS, and is now reinforced (Diaz and Pinto *et al.*, 2011, Graça *et al.*, 2012, Diaz *et al.*, 2013). Choline is degraded by the gut microflora to DMA (found decreased in this cohort) and TMA (found increased in this cohort) or converted to betaine and DMG in the methionine cycle (Salek *et al.*, 2007). Choline is also a precursor of lipids' synthesis, namely for very low density lipoproteins (VLDL) through the formation of phosphorylcholine (Zhang *et al.*, 2009). Indeed, decreased choline has been found in serum of T2DM subjects, this variation having been related to changes in lipids metabolism and gut microflora (Zhang *et al.*, 2009). The *pre-diagnosis* GDM metabolic signature was also characterized by increased NMND and decreased allantoin, the first having been previously found increased in similar GDM reports, although considering smaller cohorts (Diaz and Pinto *et al.*, 2011, Diaz *et al.*, 2013). Moreover, the same variation (increased NMND, decreased allantoin) had been reported in urine of T2DM subjects, these being suggestive of profound perturbation in nucleotide metabolism. NMND is involved in the tryptophan- NAD^+ pathway, which supplies nucleotides to the liver, being further metabolized to *N*-methyl-2-pyridone-5-

carboximide (2-Py), whilst allantoin is a degradation product of purines metabolism (Salek *et al.*, 2007). These findings suggest that a nucleotide metabolism perturbation precedes the clinical manifestation of GDM, although no variation regarding this pathway was found at the time of diagnosis. Other relevant changes include increase 3-HIVA, a metabolite previously found in one exploratory report (Diaz and Pinto *et al.*, 2011) possibly reflecting altered biotin status.

5.3 Impact of GDM treatment on urine metabolic composition

5.3.1 Unveiling urinary metabolic signatures of GDM treatment responses

For this study, maternal urine was collected from GDM affected women, during the course of treatment (either through diet alone, or diet and insulin) here named *Treated GDM* in order to gain a deeper understanding of the general metabolic response to GDM treatment. As the *Treated GDM* group included samples collected in the 2nd ($n=10$) and 3rd trimesters ($n=28$), with GA range of 17-40 g.w., a GA-matched control group was defined, here named *control group 3* ($n=16$ and $n=24$ for 2nd and 3rd trimesters respectively), with GA range of 16-39 g.w.. As was presented in section 5.1 for *control group 1*, a control group matched, as much as possible, in number and GA was selected for this comparison in order to minimize a possible confounding contribution of GA (detailed description of sample numbers and metadata in Section 2.1.2 c and Table 2.4 of Experimental Section).

The PLS-DA of the whole ^1H NMR spectra of *Treated GDM* (including both types of treatments) and controls showed a separation trend which however was not validated by MCCV. Thus, variable selection was applied and the original matrix was reduced. The resulting PLS-DA revealed a clear separation between *Treated GDM* and controls (Figure 5.5a), the model having high predictive power (see Q^2 distribution) and robustness (CR 89%, sensitivity 88% and specificity 89%), as shown in Figure 5.5b,c and summarized in Table A-III.3, Annex III. This finding indicates that, in spite of GDM treatment, maternal metabolism does not completely return to a normal situation as differences are still detectable in relation to the control group (healthy situation). Thus, the urinary metabolic signature of *Treated GDM* (Table 5.3) comprises different contents in 21 metabolites/regions. Among these, some changes were not observed in either of the

previous disease stages, seeming therefore to be specific of the effects of treatment (Figure 5.7, page 199). These comprised isoleucine, valine, malonate, *cis*-aconitate, *scyllo*-inositol, trigonelline, ascorbate, formate, lactose, mannose, and 6 new unassigned.

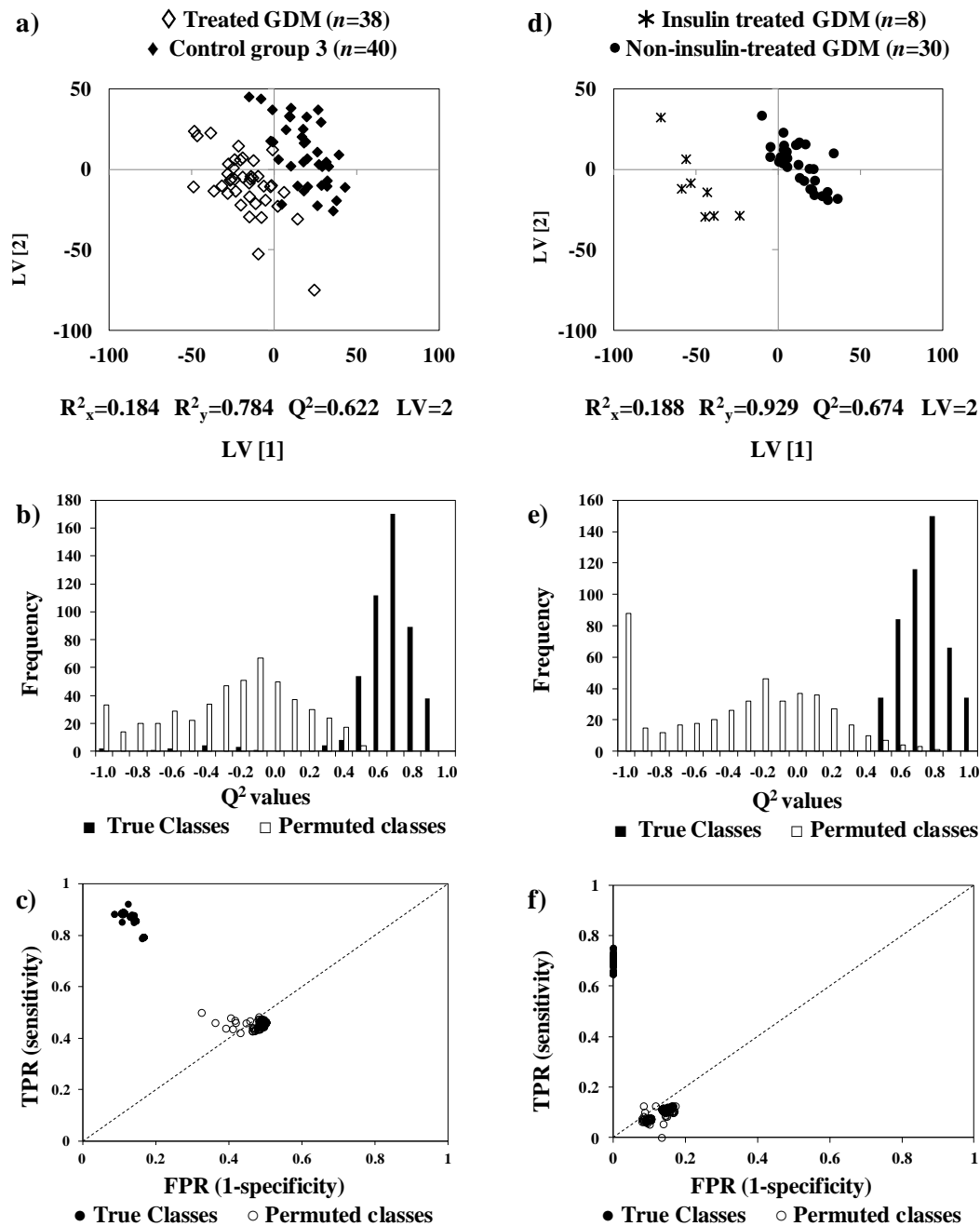


Figure 5.5: a,d) PLS-DA score plots, b,e) Q^2 and c),f) ROC plots of true and permuted models obtained for *Treated GDM* ($n=38$) vs. *control group 3* ($n=40$) (left) and for *Insulin-treated* ($n=8$) vs. *Non-insulin-treated* GDM ($n=30$) (right). TPR: true positive rate, FPR: false positive rate.

Table 5.3: Metabolite/resonance changes in maternal urine of *Treated GDM* compared to *control group 3* (left) and of *Insulin-treated vs. Non-insulin-treated* (right). ^a Chemical shifts selected by variables selection; TOCSY or STOCYSY correlated peaks in square brackets; s: singlet, d: doublet, t: triplet, q: quartet, dd: doublet of doublets, m: multiplet, br: broad. ^b Peaks possibly related to diet and/or gut microflora. Ui: unassigned compound i following numbering in Table 5.1 and Table 5.2. * Insulin-resistant changes. 2-HIBA: 2-hydroxyisobutyrate, other abbreviations defined in Table 5.1 and Table 5.2. Only *p*-values<0.05 are shown.

Treated GDM (n=38) vs. Control group 3 (n=40)			Insulin-treated (n=8) vs. Non-insulin-treated (n=30)		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)	Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
4-DEA	1.11 d	$\uparrow(0.74, 8.45 \times 10^{-4})$	2-HIBA ^b	1.36 s	$\uparrow(0.56, 9.53 \times 10^{-3})$
4-DTA	1.23 d	$\uparrow(0.84, 2.75 \times 10^{-4})$	3-aminoisobutyrate ^c	1.19 d	$\downarrow(-0.85)$
4-HPA ^b	6.86 d, 7.17 d	$\downarrow(-0.49, 1.95 \times 10^{-2})$	3-Me-histidine ^b	3.75 s, 7.78 s	$\downarrow(-0.80, 4.96 \times 10^{-2})$
4-OH-hippurate ^b	6.98 d, 7.76 d	$\downarrow(-0.12)$	4-OH-hippurate ^{b, *}	6.98 d	$\downarrow(-0.61, 3.43 \times 10^{-2})$
Acetate	1.93 s	$\uparrow(0.38)$	Choline ^b	3.20 s	$\downarrow(-0.31)$
Ascorbate	4.52 d	$\downarrow(-0.27)$	Cis-aconitate	3.12	$\downarrow(-0.17)$
cis-aconitate	3.12 d	$\uparrow(0.33)$	Citrate [*]	2.54 d, 2.69 d	$\uparrow(0.84, 7.48 \times 10^{-3})$
Citrate	2.54 d, 2.69 d	$\uparrow(0.46)$	Creatinine ^b	3.05 s, 4.06 s	$\uparrow(0.44)$
DMG	2.93 s	$\uparrow(0.48)$	DMG	2.93 s	$\downarrow(-0.65)$
Formate	8.46 s	$\uparrow(0.21)$	Furoylglycine ^{b, *}	6.65 q	$\downarrow(-0.48)$
Glutamine	2.15 m	$\uparrow(0.08)$	GAA	3.80 s	$\uparrow(0.36)$
Isoleucine	1.01 d	$\uparrow(0.48, 1.70 \times 10^{-2})$	Glutamine [*]	2.15 m, 2.47	$\uparrow(0.47)$
Lactose	4.46 d	$\uparrow(0.36)$	Glucose	3.23 dd, 3.44	$\downarrow(-0.42)$
Malonate	3.11 s	$\downarrow(-0.49)$		3.50 t, 3.72 dd	
Mannose	5.19 d	$\uparrow(0.38)$	Glycine	3.57 s	$\downarrow(-0.52)$
Scyllo-inositol ^b	3.37 s	$\downarrow(-0.35)$	Hypoxanthine	8.19 s, 8.21 s	$\uparrow(0.41)$
Sucrose ^b	5.41 d	$\downarrow(-0.51, 5.44 \times 10^{-3})$	Malonate	3.11 s	$\uparrow(0.49)$
Threonine	4.26 dd, 3.61 d	$\uparrow(0.52, 1.61 \times 10^{-2})$	NMND	8.90 d, 8.97 d,	$\downarrow(-1.08, 1.35 \times 10^{-2})$
Trigonelline ^b	4.44 s, 8.84 br,	$\uparrow(0.14)$		9.28 s	
	9.12 s		PAG ^b	7.36 t, 7.43 t	$\uparrow(0.25)$
Valine	0.99 d	$\uparrow(0.47, 6.06 \times 10^{-2})$	Tyrosine	6.90 d, 7.20 d	$\uparrow(0.34)$
Xylose ^b	3.53 dd, 4.59 d,	$\uparrow(0.58, 3.31 \times 10^{-2})$	Valine	0.99 d, 1.04 d	$\uparrow(0.10)$
	5.21 d		U15 [*]	2.12 d	$\uparrow(0.71, 3.43 \times 10^{-2})$
U10	6.48 d [6.95 d]	$\downarrow(-0.68, 1.61 \times 10^{-2})$	Unassigned spectral regions		
U11	2.84 s	$\uparrow(0.49, 6.06 \times 10^{-2})$	1.65-1.68		$\uparrow(0.47, 1.07 \times 10^{-2})$
U12	2.05 s	$\uparrow(0.59, 7.92 \times 10^{-3})$	2.82-2.83		$\uparrow(0.50, 3.77 \times 10^{-2})$
U13	2.06 s	$\uparrow(0.52, 1.30 \times 10^{-2})$	5.47-5.50		$\uparrow(0.98, 8.46 \times 10^{-3})$
U14	1.25 d [3.88]	$\uparrow(0.45, 7.59 \times 10^{-2})$	6.75-6.80		$\uparrow(0.31, 4.14 \times 10^{-2})$
U15	2.12 d	$\uparrow(1.51, 5.88 \times 10^{-7})$	+ 10 Unassigned spin systems and 13 spectral regions with <i>p</i> >0.05		
Unassigned spectral regions					
0.95-0.96		$\uparrow(0.48, 6.35 \times 10^{-2})$			
1.13-1.17		$\uparrow(0.74, 1.28 \times 10^{-3})$			
1.41-1.43		$\uparrow(0.69, 3.43 \times 10^{-3})$			
1.57-1.58		$\uparrow(0.53, 3.15 \times 10^{-2})$			
1.59-1.63		$\uparrow(0.61, 3.00 \times 10^{-3})$			
1.64-1.66		$\uparrow(0.75, 1.10 \times 10^{-3})$			
1.85-1.89		$\uparrow(0.98, 1.21 \times 10^{-4})$			
2.08-2.10		$\uparrow(0.54, 1.75 \times 10^{-2})$			
3.38-3.39		$\downarrow(-0.27, 4.91 \times 10^{-2})$			
3.58-3.60		$\downarrow(-0.36, 5.15 \times 10^{-2})$			
3.63-3.64		$\uparrow(0.50, 1.52 \times 10^{-2})$			
6.53-6.57		$\downarrow(-0.70, 2.49 \times 10^{-2})$			
6.60-6.64		$\downarrow(-0.60, 2.05 \times 10^{-3})$			
8.48-8.51		$\downarrow(-0.54, 2.76 \times 10^{-2})$			
8.57-8.59		$\downarrow(-0.59, 1.61 \times 10^{-2})$			
9.36-9.37		$\downarrow(-0.60, 5.26 \times 10^{-3})$			
+ 6 unassigned spin systems and 12 spectral regions with <i>p</i> >0.05					

The specific metabolic effect of insulin treatment compared to dietary control was sought, in order to unveil specific urinary signatures of each therapeutic approach (insulin and diet or diet alone). For this study, urine samples from the *Treated GDM* group were divided into *Insulin-treated* ($n=8$) and *Non-insulin-treated* ($n=30$, i.e. diet only), and compared through PLS-DA (after variable selection). A significant impact on maternal urine was found in relation to the two types of GDM treatments, as shown in the PLS-DA score plot in Figure 5.5d. MCCV confirmed the high predictive power (seen by Q^2 distribution) and robustness (CR 94%, sensitivity 100% and specificity 93%) of this model, as shown in Figure 5.5e,f and Table A-III.3.

The *Insulin-treated* group showed higher contents of 2-hydroxyisobutyrate (2-HIBA), citrate, creatinine, GAA, glutamine, hypoxanthine, malonate, PAG, tyrosine, valine, an unassigned doublet at 2.12 ppm (U15) and 4 unassigned spectral regions with $p<0.05$, and lower contents of 3-aminoisobutyrate, 3-me-histidine, 4-OH-hippurate, choline, *cis*-aconitate, DMG, furoylglycine, glucose, glycine, and NMND compared to the *Non-insulin-treated* group (diet alone) (Table 5.3, right). Variations were also analyzed in terms of responsiveness and resistance to insulin treatment, i.e. variations common in nature and direction to the *Diagnosed GDM* state (present prior to- and during treatment) were considered *insulin-resistant*. On the other hand, metabolites found to vary upon insulin treatment but not characterizing the *Diagnosed GDM* group were considered *insulin-responsive*. Within the insulin-treatment signature here identified, 4-OH-hippurate, citrate, furoylglycine and glutamine (noted with * in Table 5.3, right) were found to be insulin-resistant while the remaining variations were identified as insulin-responsive.

Finally, the concomitant comparison of *Insulin-treated* vs. *Non-insulin-treated* vs. *control group 3* was carried out, aiming at understanding the efficacy of each treatment in returning to normal metabolism (characteristic of controls). This was computed by a 3-class PLS-DA (Figure 5.6) which showed a clear separation between controls, subjects treated by insulin plus diet and those treated by diet alone, thus confirming different metabolic responses to the two therapeutic approaches.

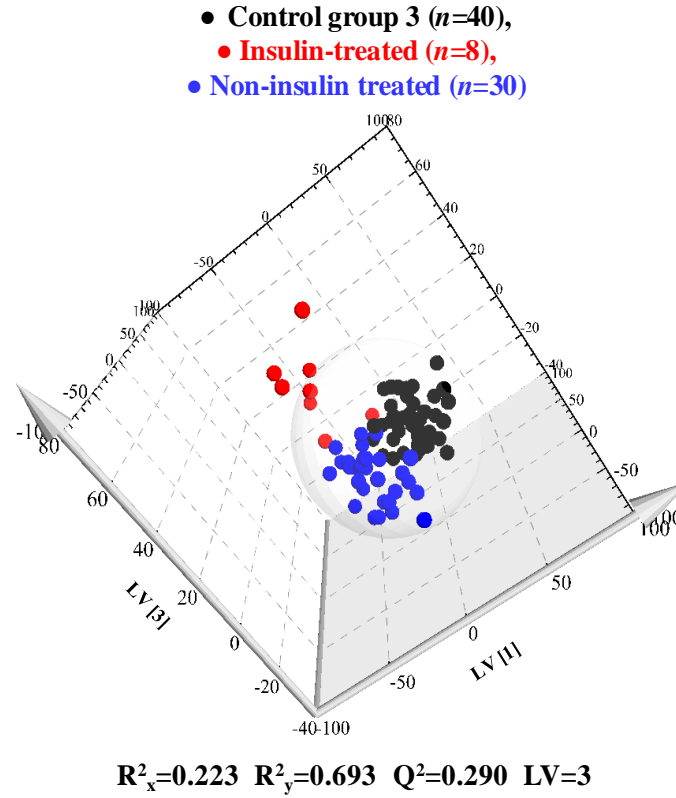


Figure 5.6: 3LV PLS-DA score plot of control group 3 vs. Insulin-treated GDM vs. Non-Insulin-treated GDM.

The class separation seen in Figure 5.6 shows that *Insulin-treated* subjects are more different from controls than *Non-insulin-treated* women, as expressed by the higher distance in the score plot relatively to the control group. This result confirms that maternal metabolism does not completely recover to a normal situation in either treatments although *Non-insulin-treated* subjects' metabolism seems closer to that of controls, compared to *Insulin-treated* cases. A possible explanation for this finding is that insulin therapy is prescribed to pregnant women that maintain high plasma glucose levels after initiation of the dietary intervention (specific cutoff values shown in Section 2.1.2, Experimental Section), i.e. presenting higher GDM severity. Thus, the higher differentiation found for subjects undergoing insulin treatment may in fact reflect the severity of the disease rather than the treatment efficacy itself. Finally, the results obtained with the 3-class PLS-DA model indicate the ability of the method to follow individual treatment responses, with potential application in treatment efficacy assessment and identification of treatment adjustments requirements.

5.3.2 Proposed metabolic interpretation of the effect of GDM treatments

General effects of treatment

A schematic summary of the metabolic disturbance found in this study upon GDM treatments is shown in Figure 5.7, with indication of two metabolites (4-HPA and citrate) whose levels are not improved (i.e. are not closer to controls) by treatment. Regarding the general effect of GDM treatment (*Treated GDM* group), compared to controls, a marked perturbation in amino acids metabolism was evident through the significant increase in threonine, and its catabolic products 4-DTA and 4-DEA, along with increased valine and isoleucine. Increased excretion of threonine, valine and isoleucine suggests their underuse for gluconeogenesis, as has previously been reported in relation to T2DM treatment (van Doorn *et al.*, 2007). Regarding branched chain amino acids (BCAA), increased valine and isoleucine had been found in umbilical cord plasma and serum and in maternal serum of GDM-affected pregnancies, possibly due to altered placental amino acid exchange and/or placental-fetal metabolism (Cetin *et al.*, 2005, Dani *et al.*, 2013). Higher acetate, *cis*-aconitate and citrate, precursors and intermediates of the TCA cycle, were found increased relatively to controls, although these variations were not found to be individually significant, suggesting a slightly higher TCA operation in GDM cases.

Other relevant changes were again found in metabolites deriving from diet and/or gut microflora, as was the case for *Diagnosed* and *pre-diagnosed GDM*. Accordingly, GDM treatment lead to a decrease in *scyllo*-inositol and an increase in trigonelline (no changes in hippurate were seen, contrary to previous reports showing hippurate variation in treated T2DM subjects (van Doorn *et al.*, 2007)). In addition, we found decreased 4-HPA and 4-OH-hippurate commonly to the *Diagnosed GDM* group, thus possibly reflecting a characteristic perturbation of the disease or treatment resistance, as these metabolites were left unchanged upon treatment. Changes in the urinary excretion of sugars were found in *Treated GDM*, namely increased xylose, mannose and lactose and decrease sucrose. As all subjects from the *Treated GDM* group were under strict dietary control, changes in sugars and metabolites deriving from diet and/or gut microflora can be a direct reflection of these restrictions.

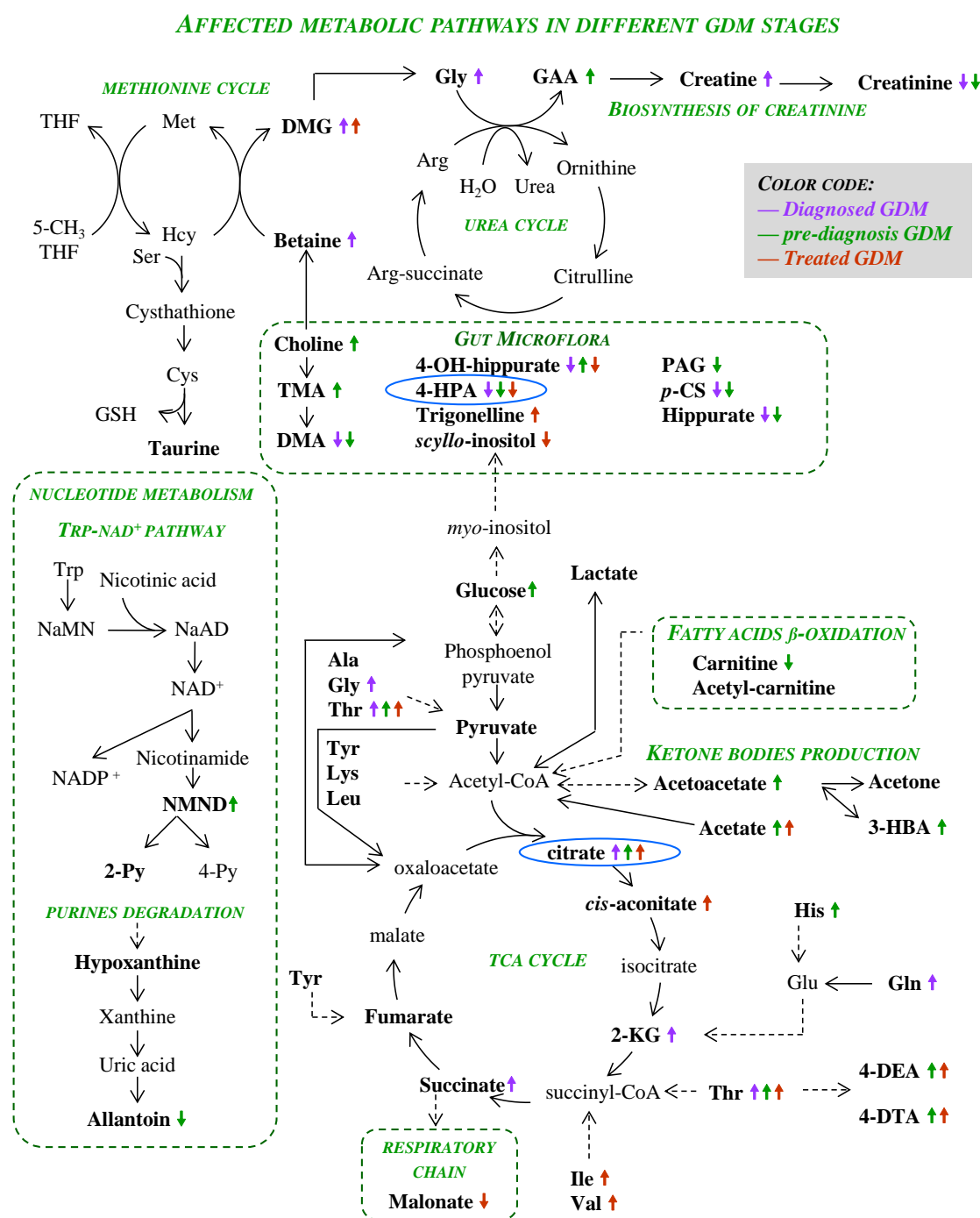


Figure 5.7: Schematic representation of the metabolic pathways possibly affected in several GDM stages. Metabolites in bold were identified in ¹H NMR spectra of maternal urine. Colored arrows indicate direction of variation in each group. Dashed arrows indicate transformations occurring through several reactions. Amino acids in 3-letter code; NaMN: nicotinic acid mono nucleotide, NaAD: nicotinic acid adenine dinucleotide, NAD⁺: nicotinamide adenine dinucleotide, NADP: nicotinamide dinucleotide phosphate, NMND: *N*-methyl-nicotinamide, 2-Py- *N*-methyl-2-pyridone-5-carboxamide, 4-Py: *N*-methyl-4-pyridone-3-carboxamide, 4-DEA: 4-deoxyerythronic acid, 4-DTA: 4-deoxythreonic acid, 3-HBA: 3-hydroxybutyrate, DMA: dimethylamine, TMA: trimethylamine, DMG: dimethylglycine, GAA: guanidoacetate, Hcy: homocysteine, GSH: glutathione, THF: tetrahydrofolate. Blue ellipses indicate metabolites which resist to treatment.

Effect of insulin and diet treatment compared to diet alone

A summary scheme of the metabolic pathways affected due to insulin treatment is shown in Figure 5.8 along with indication of insulin-resistant in red and insulin-responsive variations in blue. Regarding the metabolic effect of *Insulin treatment* (compared to the *Non-insulin-treated* group, i.e. diet alone), amino acids valine, tyrosine and glutamine were found to increase (the latter found to be insulin-resistant, meaning also present prior to treatment), possibly due to their underuse for energy metabolism, whilst decreased glycine suggests a further involvement in methionine metabolism (Figure 5.8). In fact, the concomitant decreases in DMG, and choline in the *Insulin-treated* group can be interpreted as a slowing down of the methionine cycle. This is supported by a previous report of an animal T2DM model (Ratnam *et al.*, 2006) showing that insulin inhibits betaine-homocysteine methyltransferase (BHMT), which transfers a methyl group from betaine to homocysteine forming methionine and DMG. Increased GAA and creatinine, also found in *Insulin-treated* subjects, possibly reflect a perturbation of the urea cycle (Brosnan and Brosnan, 2010) and/or creatinine biosynthesis. It has been suggested that insulin action decreases the expression of carbamoyl phosphate synthetase, a key enzyme in the urea cycle (Wilcox, 2005) and we propose that this possibly explains the deregulation of this pathway.

Urine of *Insulin-treated* subjects were also found to have lower *cis*-aconitate and higher malonate levels, possibly reflecting changes in TCA cycle and respiratory chain (Amaral *et al.*, 2012). Moreover, a significant increase in citrate was found in *Insulin-treated* women, this metabolite being found to be insulin-resistant (i.e. the high levels found prior to treatment are maintained). Citrate has been reported to increase with hyperglycemia in GDM subjects (Sachse *et al.*, 2012) and in T1DM subjects with poor glycemic control (Deja *et al.*, 2013). As mentioned above, insulin therapy is prescribed for pregnant women with more severe GDM, i.e. women that maintain high plasma glucose levels after initiation of the dietary intervention. Although information about plasma glucose levels or glycemic control was not available for this group of subjects, it seems possible that the increase in urinary citrate in the *insulin-treated* group reflects the higher severity of GDM in these women. Another finding is that of urinary glucose excretion, which was found to decrease in *insulin-treated* subjects, possibly reflective of an effective treatment action. However, as urinary glucose is not a reliable indicator of metabolic

control in GDM due to changes in renal glucose handling in pregnancy (Hod *et al.*, 2008) no definitive conclusions can be drawn.

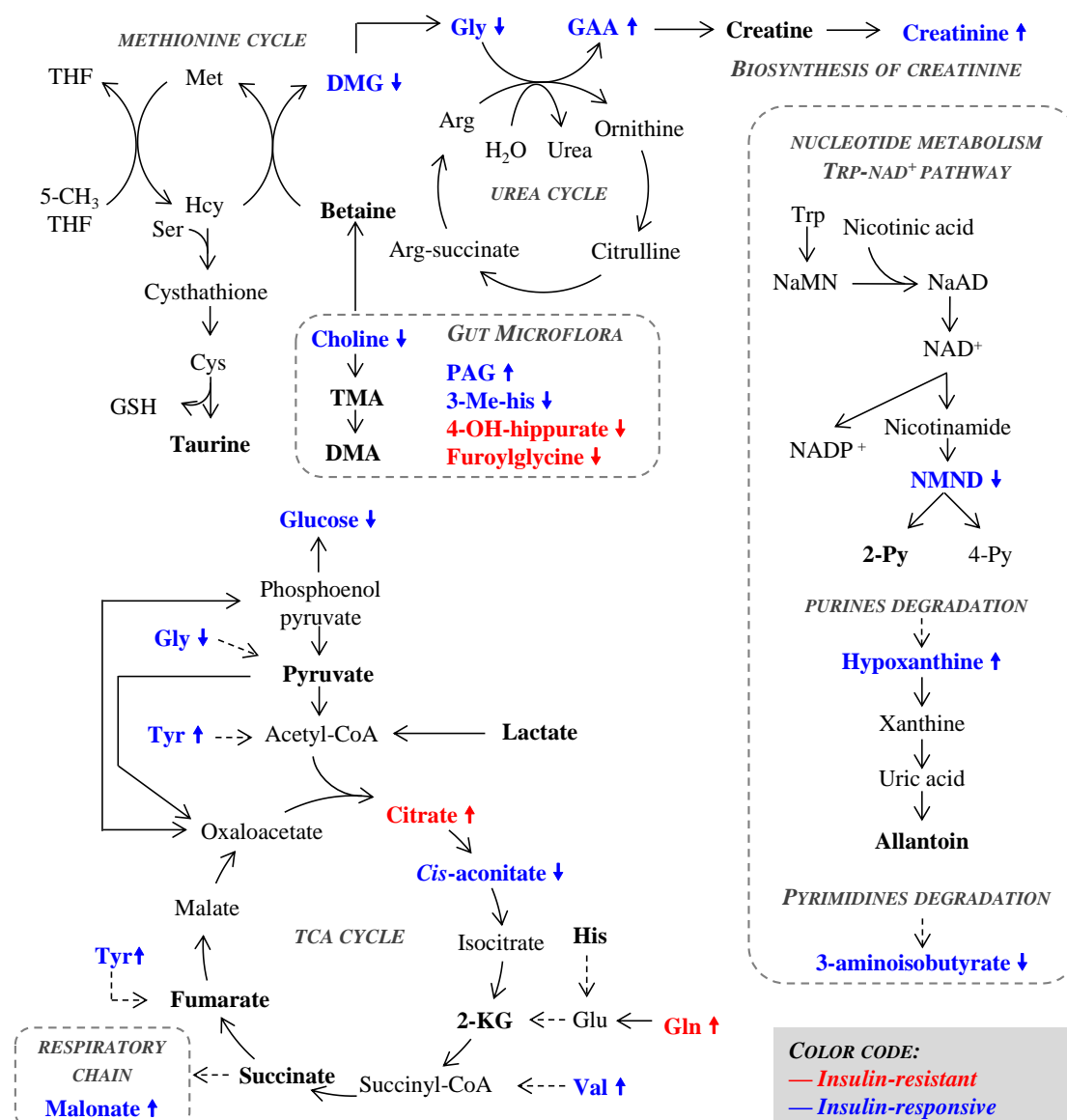


Figure 5.8: Schematic representation of the metabolic pathways possibly affected in GDM women undergoing insulin and diet treatment compared to those only under dietary treatment. Metabolites in bold were identified in ¹H NMR spectra of maternal urine. Metabolites found insulin-resistant are shown in red, those insulin-responsive are shown in blue. Short arrows indicate direction of variation in the *Insulin-treated* group. Dashed arrows indicate transformations occurring through several reactions. Amino acids in 3-letter code; Abbreviations defined in Figure 5.7

Regarding metabolites deriving from diet and/or gut microflora, higher 2-HIBA and PAG along with lesser 3-methylhistidine, furoylglycine and 4-OH-hippurate were found in women treated with insulin plus diet compared to those treated by diet alone (the

latter two metabolites being insulin-resistant). A significant decrease in NMND excretion was found in *Insulin-treated* women, this metabolite being involved in nucleotide metabolism, particularly in the tryptophan-NAD⁺ pathway. Increased urinary NMND has been found in T2DM (Salek *et al.*, 2007) and, in this work, in the *pre*-diagnosis GDM stage. Moreover, 3-aminoisobutyrate, formed through the degradation pathway of pyrimidines (van Kuilenburg *et al.*, 2004), and hypoxanthine, formed from purine degradation (Enea *et al.*, 2010), were found decreased in *Insulin-treated* subjects, again supporting a perturbation in nucleotide metabolism due to insulin therapy.

Regarding relevant unassigned spin resonances, emphasis must be given to the unassigned doublets at 1.25 ppm (U14) and at 2.12 ppm (U15). U14 was found decreased at the time of diagnosis (*Diagnosed GDM* group) and increased upon GDM treatment, irrespectively of the therapeutic approach, thus possibly reflecting a general treatment effect. U15 was found to be significantly increased upon treatment (*Treated GDM* group compared to controls), with a further increase in *Insulin-treated* subjects, possibly reflecting either disease severity and/or specific insulin treatment effect.

5.4 Prediction of GDM treatment response: a possible pharmacometabolomic application

This study aimed at exploring the possibility of predicting future GDM treatment requirements based on the individual *pre-treatment* urinary profile, envisaging improved personalized therapeutic approaches. This potential application is the basis of the pharmacometabolomic concept, which “seeks to predict the response of an individual to a stimulus (e.g., drug, toxin, surgery, nutrition) *prior* to stimulus or other perturbation” (Nicholson *et al.*, 2011). Thus, the hypothesis here tested was that each individual has an inherent phenotype which correlates to her response to treatment.

5.4.1 Unveiling a *pre-treatment* metabolic phenotype predictive of GDM treatment requirements

In spite of the low number of samples available, the possibility to identify a potentially predictive metabolic phenotype of future insulin treatment requirements was attempted by comparing the subgroups (found within the *Diagnosed GDM* group) of

subjects *later* requiring insulin treatment ($n=6$) and subjects not requiring insulin treatment ($n=7$), all samples being collected prior to treatment initiation.

The PLS-DA of the original ^1H NMR dataset revealed a separation trend in the score plot (not shown), although the model was not found to have predictive power or robustness (as shown by MCCV parameters in Table A-III.3, Annex III). Thus, upon the application of variable selection the reduced dataset was resubmitted to PLS-DA and MCCV, the resultant model having CR 94%, sensitivity 95% and specificity 94%. The corresponding PLS-DA model score plot, Q^2 distribution and ROC are shown in Figure 5.9 and Table A-III.3.

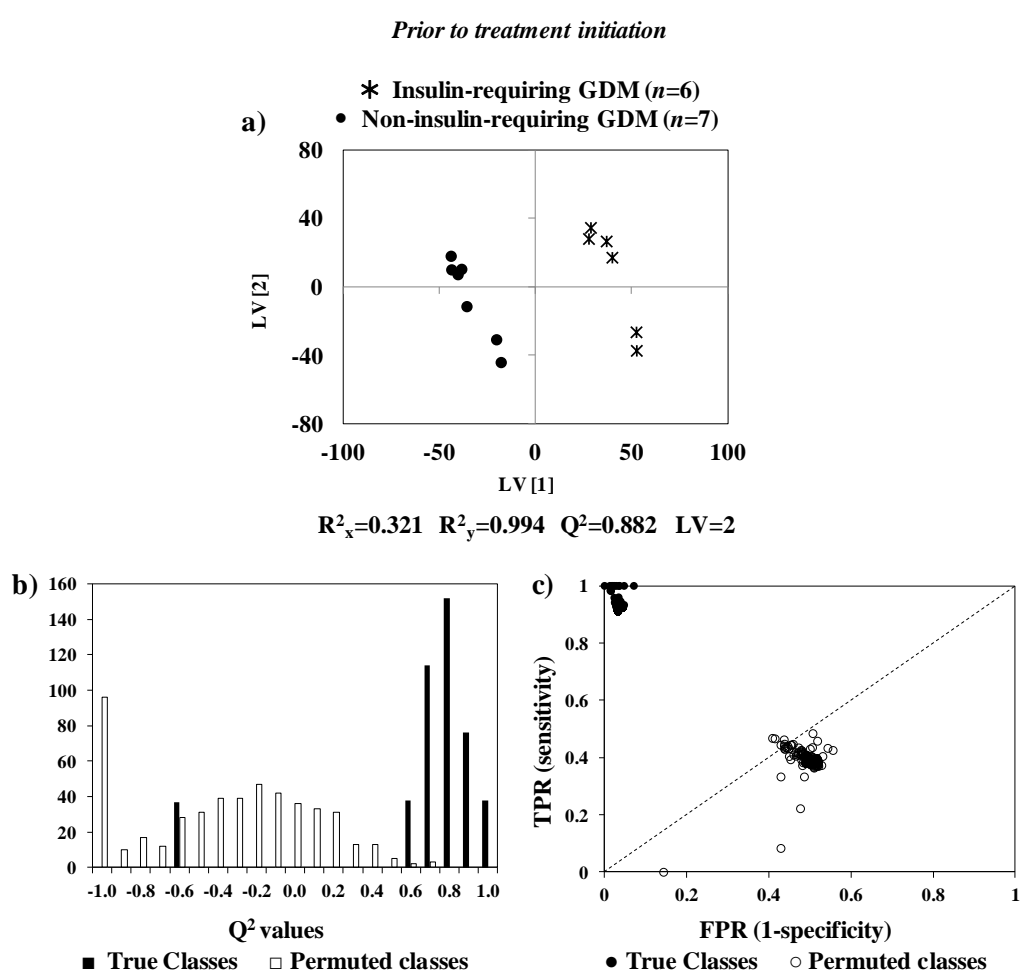


Figure 5.9: a) PLS-DA score plots, b) Q^2 and c) ROC plots of true and permuted models obtained for *Insulin-requiring* GDM ($n=6$) vs. *Non-insulin-requiring* GDM ($n=7$). TPR: true positive rate, FPR: false positive rate.

The subset of variables selected, with basis on PLS-DA separation, unveiled a metabolic signature potentially predictive of future insulin treatment requirement (Table

5.4). This signature included variations in 23 metabolites, 1 unassigned singlet (at 0.63 ppm) and 2 spectral regions with $p < 0.05$. This means that subjects *later* needing insulin have an original profile characterized by higher 2-Py, 4-HPA, acetate, betaine, creatine, creatinine, hippurate, lysine, NMND and tyrosine and lower 2-KG, 3-aminoisobutyrate, 3-HBA, 3-HIVA, 4-OH-hippurate, acetoacetate, acetone, *cis*-aconitate, citrate, formate, lactose, malonate, succinate and unassigned singlet at 0.63 ppm (U16). Moreover,

Table 5.4: Metabolite/resonance changes in maternal urine of *Insulin-requiring* compared to *Non-insulin-requiring* subjects. ^a Chemical shifts selected by variables selection;; s: singlet, d: doublet, t: triplet, q: quartet, dd: doublet of doublets, m: multiplet, br: broad. ^b Peaks possibly related to diet and/or gut microflora. *Insulin-dependent changes. Ui: unassigned compound i following numbering in Table 5.1, Table 5.2 and Table 5.3. 2-KG: 2-ketoglutarate, 2-Py: *N*-methyl-2-pyridone-5-carboxamide, 3-HBA: 3-hydroxybutyrate, 3-HIVA: 3-hydroxyisovalerate, 4-HPA: 4-hydroxyphenylacetate, NMND: *N*-methyl-nicotinamide. Only p -values < 0.05 are shown.

<i>Insulin-requiring (n=6) vs. Non-insulin-requiring (n=7)</i>		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, p -value)
2-KG*	2.45 t, 3.01 t	↓(-0.59)
2-Py*	6.67 d	↑(0.03)
3-aminoisobutyrate*	1.19 d	↓(-0.76)
3-HBA*	1.20 d	↓(-1.22)
3-HIVA*	1.27 s, 2.37 s	↑(1.01)
4-HPA ^{b, *}	7.17 d	↑(1.08)
4-OH-hippurate ^b	3.95 s, 6.98 d, 7.76 d	↓(-1.77, 1.4×10^{-2})
Acetate*	1.93 s	↑(0.23)
Acetoacetate*	2.28 s	↓(-0.85)
Acetone*	2.23 s	↓(-1.00)
Betaine	3.27 s	↑(0.80)
<i>cis</i> -aconitate*	3.12 d	↓(-0.18)
Citrate*	2.54 d	↓(-0.14)
Creatine ^{b, *}	3.04 s	↑(0.27)
Creatinine ^b	3.05 s	↑(0.45)
Formate*	8.46 s	↓(-0.97)
Hippurate ^{b, *}	8.52 br	↑(1.08)
Lactose*	4.46 d	↓(-0.46)
Lysine*	1.73 m	↑(0.50)
Malonate*	3.11 s	↓(-0.46)
NMND*	4.48 s	↑(0.19)
Succinate*	2.41 s	↓(-0.80)
Tyrosine*	7.19 d	↑(0.77)
U16*	0.63 s	↓(-1.45, 3.5×10^{-2})
Unassigned spectral regions		
1.41-1.44		↓(-1.65, 1.40×10^{-2})
5.44-5.45		↑(0.62, 2.21×10^{-2})
+ 17 unassigned spin systems and 34 spectral regions with $p > 0.05$		

metabolites were analyzed in terms of insulin dependence, meaning that variations found different, in nature and direction, compared to the *Diagnosed GDM* stage were classified as insulin-dependent (i.e. only present in women later requiring insulin treatment). The remaining variations (namely betaine, creatine and 4-OH-hippurate, noted with * in Table 5.4) were found commonly, in nature and direction, with the *Diagnosed GDM* group, thus considered non-insulin-dependent.

In order to evaluate if this *pre-treatment* urinary signature distinguishes future treatment requirements, not only from each other but also from controls (no disease), a PLS-DA was computed considering simultaneously the three groups: *control group 1* vs. *Insulin-requiring* vs. *Non-insulin-requiring* subjects. The resulting 3LV PLS-DA score plot (Figure 5.10) showed a clear separation between the three groups with high predictive power (Q^2 0.439), suggesting that each of the three groups hold a specific urinary signature.

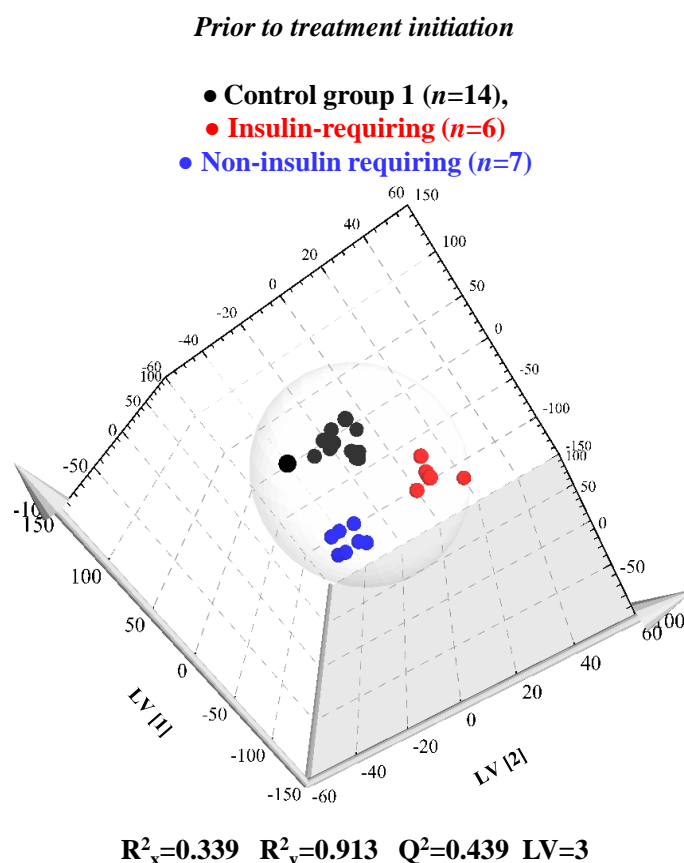


Figure 5.10: 3LV PLS-DA score plot of *control group 1* vs. *Insulin-requiring* GDM vs. *Non-insulin-requiring* GDM.

This result reinforces the value of maternal urine for a pharmacometabolomic application, which can be implemented towards the development of early and personalized therapeutic approaches.

5.4.2 Proposed metabolic interpretation of *pre*-treatment urinary profiles

The individual urinary phenotype predictive of future insulin therapy requirements is schematized in Figure 5.11, with indication of insulin-dependent variations in red and non-insulin-dependent variations in blue. The results suggest that subjects later requiring insulin treatment show *a priori* perturbations in energy metabolism namely through the increase in lysine and tyrosine, both precursors of acetyl-CoA, and fumarate. An increase in TCA precursor acetate and decrease in its intermediates (2-KG, *cis*-aconitate, citrate, and succinate) along with lower ketone bodies excretion (3-HBA, acetoacetate and acetone) were also found. The overall findings are consistent with a slowing down of TCA cycle operation and ketone bodies production leading to a decrease in acetyl-CoA production and accumulation of its precursors. Still in connection with the TCA cycle, the decrease in malonate, a competitive inhibitor of succinate in the respiratory chain, possibly suggests perturbations in the respiratory chain and/or a consequence of succinate's decrease. Regarding metabolites deriving from the gut microflora, decreased 4-OH-hippurate and increased 4-HPA and hippurate, were found in the *Insulin-requiring* group. An involvement of nucleotide metabolism was also registered in women further requiring insulin therapy, as suggested by the tandem increase of NMND and 2-Py (despite the latter having a residual change) and decreased 3-aminoisobutyrate, a product of pyrimidine degradation (van Kuilenburg *et al.*, 2004). These findings hint at a perturbation in nucleotides metabolism, in particular the tryptophan-NAD⁺ pathway, in women later requiring insulin therapy, thus possibly reflecting disease severity. Increased urinary 3-HIVA was also found in *Insulin-requiring* women, this metabolite having been found in urine in connection to cadmium exposure and related with mitochondrial damage (Ellis *et al.*, 2012) whilst in connection to GDM, the opposite variation was found in the *pre*-diagnosis GDM stage (Section 5.2). Increased creatinine and creatine were also found for *Insulin-requiring* women, the latter having also been found at diagnosis, possibly reflective of the disease and not insulin requirement.

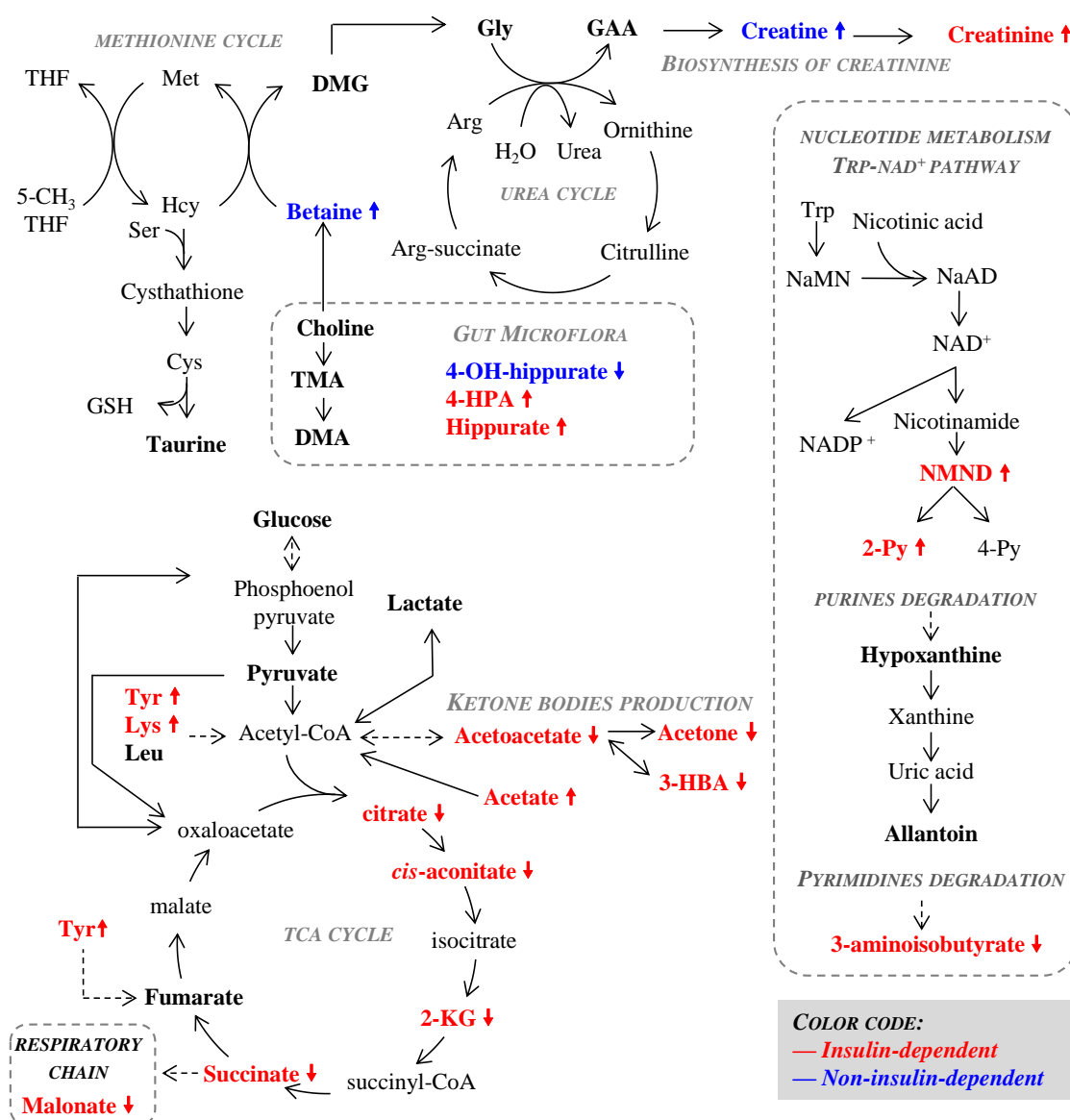


Figure 5.11: Schematic representation of the metabolic pathways possibly affected in GDM women further requiring insulin treatment compared that never required insulin therapy. Metabolites in bold were identified in ¹H NMR spectra of maternal urine. Metabolites found insulin-dependent are shown in red, those non-insulin-dependent are shown in blue. Short arrows indicate direction of variation in the *Insuli- requiring* group (relatively to *Non-insulin-requiring*). Dashed arrows indicate transformations occurring through several reactions. Amino acids in 3-letter code; Abbreviations defined in Figure 5.7.

5.5 Summary of the impact of GDM in maternal urine

The results presented in this chapter demonstrate the impact of three GDM stages (*diagnosed, pre-diagnosis* and *treated GDM*) in maternal urine compared to GA-matched controls, with the main biochemical findings and validation results summarized in Figure 5.12.

Group studied	Main biochemical characteristics	Validation parameters
<i>Diagnosed</i> GDM	TCA cycle, amino acids and methionine metabolism, renal function and gut microflora	CR: 96% Sensitivity: 93% Specificity: 100%
<i>Pre-diagnosed</i> GDM	Hyperglycemia and increased citrate, ketone bodies production and fatty acids β -oxidation, amino acids and nucleotide metabolism, biotin status, gut microflora	CR: 87% Sensitivity: 84% Specificity: 89%
<i>Treated</i> GDM (insulin and/or diet)	TCA cycle, amino acids metabolism and gut microflora	CR: 89% Sensitivity: 88% Specificity: 89%
<i>Insulin-treated</i> GDM	TCA cycle, amino acids, methionine metabolism and nucleotides metabolism, urea cycle and gut microflora	CR: 94% Sensitivity: 100% Specificity: 93%
<i>Insulin-requiring</i> GDM	TCA cycle, amino acids and nucleotides metabolism, respiratory chain, renal function, and gut microflora	CR: 94% Sensitivity: 95% Specificity: 94%

Figure 5.12: Summary of the main biochemical findings suggested for each GDM stage and corresponding validation parameters.

The first novelty presented in this chapter, with high potential value in the clinical environment, was the possibility of identifying women later developing GDM prior to any clinical manifestation of the disease (following two previous publications resultant of this work). It was here found that 2nd trimester maternal urine carries a metabolic signature predictive of GDM development, with CR 87%, sensitivity 84% and specificity 89%. The lower model validation parameter obtained was the sensitivity or true positive rate (TPR), i.e. meaning that some women are still not being identified as *later* developing GDM. Higher number of samples and external validation are still required to validate a possible predictive method for GDM based on urine metabolomics.

The second novelty presented in this chapter was the study of the general effect of GDM treatment and the demonstration that maternal metabolism does not return completely to the normal/healthy situation. Whithin the registered changes, only renal function and methionine metabolism appear to be normalized. Additionally, it was found that subjects treated by insulin plus diet compared to diet alone hold a particular metabolic signature (CR 89%, sensitivity 88% and specificity 89%) and that women treated only by

diet were more similar to controls than those treated by insulin and diet (possibly due to different disease severity). Moreover, *Insulin-treated* subjects were differentiated from *Non-insulin-treated* subjects with CR 94%, sensitivity 100% (meaning that all insulin-treated women are correctly classified) and specificity 93%. The method here presented may provide a valuable tool to assess GDM treatment efficacy and treatment adjustment requirements.

The final part of this work explored the novel possibility of predicting individual future treatment requirements, i.e. a possible pharmacometabolomic application of maternal urine. The results here presented showed that women *later requiring insulin treatment* were distinguished from those never requiring insulin treatment with CR 95%, sensitivity 94% and specificity 95%. In spite of the exploratory nature of this model, given the very limited number of samples available for this study ($n=6$ and $n=7$), these results support the premise of predicting individual treatments requirements towards the development of personalized therapeutic approaches. Confirmations of these findings are still required through the use of larger cohorts and external validation.

6. Newborn health status seen through urine metabolomics

This chapter describes the results obtained for the NMR metabolomic study of newborn urine to characterize a) the metabolic phenomena accompanying newborn disorders and b) the impact of prenatal disorders on newborns health, in the first days of life. These findings are presented in four sections, the first section (6.1) encompassing the characterization of newborn urine composition. The following section (6.2) presents the results obtained for the impact of potential confounding effects, namely gender, delivery mode, day of life (day 1 and day 2) and gestational age at birth (after 37 gestational weeks, g.w.). The third section (6.3) shows the study of the effects of newborn disorders on newborn urine, namely addressing prematurity (birth before 37 g.w.), respiratory depression (newborns requiring oxygen supply after birth), large for gestational age (LGA, birth weight above 90th percentile), malformations (comprising cardiac, renal, soft tissues, malformations and situs inversus), *pre*-jaundice (babies later developing hyperbilirubinemia requiring phototherapy) and PROM (membrane rupture prior to labor after 37th g.w.). The above are presented and discussed in decreasing order of impact on newborn's urine. The following section 6.4 aims at understanding the impact of the *in utero* period in newborn's health, specifically the effect of being born to mothers affected by GDM, respiratory disease, thyroid disease, chronic hypertension and psychiatric disorders. The concomitant study of confounding effects and several newborns and maternal disorders enabled their comparison for the unraveling of disease-specific changes, this aspect being discussed in the final subchapter 6.5.

6.1 Typical NMR urine spectra of newborn urine and assignments

Contrary to an initial idea that newborn urine might be simpler than adult urine, it was observed that the ¹H NMR spectrum of newborn urine is as complex as adult urine. The average ¹H NMR spectrum of the control group is shown in Figure 6.1a along with some assignments. There are major differences in the composition of newborn urine and that of pregnant and non-pregnant women (previously shown in Figure 3.1). For instance, the aromatic region of urine spectrum characterized by metabolites deriving from the gut microflora, has significantly lower intensity in newborns (compared to adult pregnant and non-pregnant women), which is consistent with a lower microbiota colonization

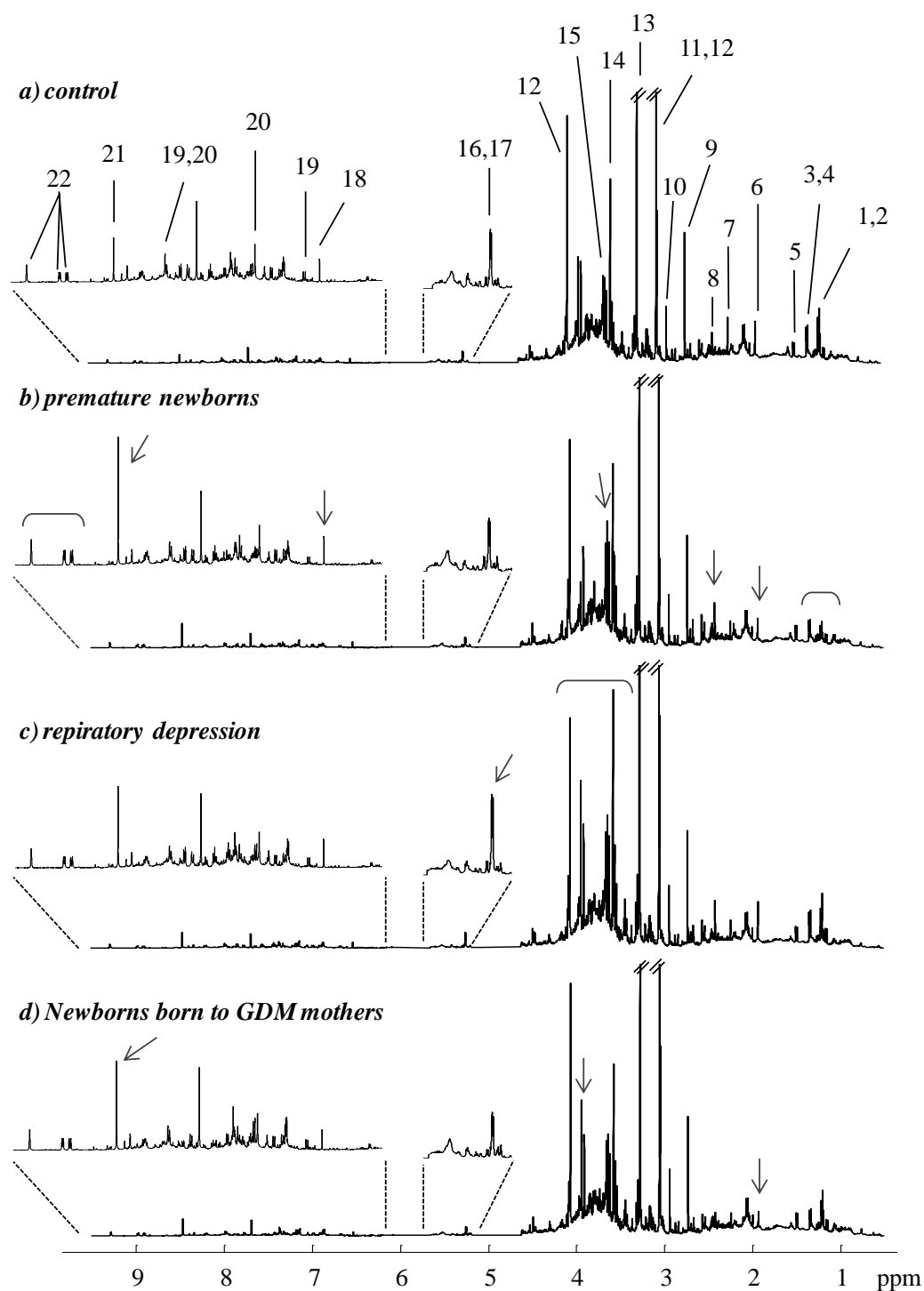


Figure 6.1: Average 500 MHz ^1H NMR spectra of a) healthy newborns (control group), and newborns born b) premature, c) affected by a respiratory depression episode after birth and d) newborns born to GDM mothers groups. Some assignments are noted on a) and major visual changes (see arrows) are indicated in b,c and d) for each disorder group. Legend: 1: 3-aminoisobutyrate, 2: 3-HBA, 3: lactate, 4: threonine, 5: alanine, 6: acetate, 7: acetone, 8: succinate, 9: DMA, 10: DMG, 11: creatine, 12: creatinine, 13: betaine, 14: glycine, 15: *myo*-inositol, 16: glucose, 17, lactose. 18: fumarate, 19: 2-Py, 20: histidine, 21: formate, 22: NMND. Abbreviations defined in Table 6.1.

and activity upon the first days of life. It is known that the newborns' microbiota is seeded at birth and influenced by the delivery mode (vaginal vs. cesarean section), breast feeding and weaning and that the microflora is highly developed over the first weeks of life (Nicholson *et al.*, 2012), thus justifying the lower intensity of the aromatic region found in newborn urine in this study.

In total, 56 metabolites were identified in newborn urine, including amino acids and derivatives, organic acids, and sugars. Within the resulting list of assignments only phenylacetylglutamine (PAG) and indoxyl sulphate (IS) were here identified in newborn urine, for the first time to our knowledge (noted with * in Table 6.1) Assignments were carried out with basis on 2D NMR experiments (TOCSY and HSQC), STOCYSY and comparison with Bruker Biorefcode 2.0.0 (courtesy of Bruker BioSpin, Rheinsteten, Germany) and HMDB (Wishart *et al.*, 2009) databases, as previously described in Chapter 3.1. The complete list of assignment is show in Table 6.1, along with the corresponding ^1H chemical shifts, multiplicity, assignment and, when possible, ^{13}C chemical shifts. Metabolite abbreviations are defined in the same table and used throughout this section.

Table 6.1: List of assignments in newborn urine ^1H NMR spectra, at pH 7.00 ± 0.02 . Assignments carried out with basis on literature, spectral databases Bruker Biorefcode and HMDB (Wishart *et al.*, 2009), as well as 2D NMR experiments (TOCSY and HSQC), and spectra of standards. s: singlet, d: doublet, t: triplet, q: quartet, dd: doublet of doublets, m: multiplet. *Metabolites identified in newborn urine for the first time to our knowledge.

Assignment	$\delta_{\text{H ppm}}$ (multiplicity, assignment/ $\delta_{\text{C ppm}}$)
1-methylhistidine	3.07 (dd, βCH_2); 3.16 dd, $\beta'\text{CH}_2$); 3.72 (s, CH_3); 3.96 (dd, αCH_2); 7.05 (s, ring); 7.78 (s, ring)
2-hydroxyisobutyrate (2-HIBA)	1.36 (s, CH_3)
2-ketoglutarate (2-KG)	3.45 (t, βCH_2), 3.1 (t, γCH_2)
3-aminoisobutyric acid	1.19 (d, $\text{CH}_3/17.87$); 2.61 (m, αCH); 3.06 (dd, βCH_2)
3-hydroxybutyrate (3-HBA)	1.20(d, CH_3); 2.31(m, CH_2); 2.41 (m, CH_2); 4.15 (m, CH)
3-hydroxyisovaleric acid (3-HIVA)	1.27 (s, $\beta\text{CH}_3/31.02$); 2.37 (s, αCH_2)
3-methylhistidine	3.28 (dd, βCH_2); 3.75 (s, NCH_3); 3.97(dd, αCH); 7.15 (s, C6H); 8.12 (s, C2H)
4-deoxyerythronic acid (4-DEA)	1.11 (d, $\gamma\text{CH}_3/18.26$); 4.08 (d, $\alpha\text{CH}/78.67$); 4.10 (m, $\beta\text{CH}/71.54$)
4-deoxythreonic acid (4-DTA)	1.23 (d, $\gamma\text{CH}_3/21.37$); 3.84 (d, $\alpha\text{CH}/79.04$); 4.12 (m, $\beta\text{CH}/71.59$)
4-hydroxyhippurate	3.95 (s, CH_2); 6.98 (d, C3H, C5H ring); 7.76 (d, C4,2H, C6H ring)
4-hydroxyphenylacetate (4-HPA)	3.46 (s, CH_2); 6.86 (d, C3H, C5H ring); 7.17 (d, C2H, C6H ring)
Acetate	1.93 (βCH_3)
Acetoacetate	3.46 (s, CH_3); 2.29 (s, CH_2)

Assignment (<i>continues</i>)	δ_{H} ppm (multiplicity, assignment/ δ_{C} ppm)
Acetone	2.23 (s, CH ₃)
Adipate	1.55 (m, β CH ₂), 1.91 (m, α CH ₂)
Alanine	1.49 (d, β CH ₃ /18.99); 3.78 (q, α CH)
Allantoin	5.39 (s, CH/66.19)
Ascorbate	3.76 (m, CH ₂ (OH)); 4.01 (m, CH (OH)); 4.52 (d, C1H)
Betaine	3.26 (s, CH ₃ /56.13); 3.90 (s, CH ₂ /69.11)
Choline	3.20 (s, N(CH ₃) ₃ /56.52); 3.52 (m, NH/70.60); 4.07 (m, CH ₂ (OH))
<i>cis</i> -aconitate	3.12 (d, CH/46.13); 5.79 (t, CH ₂ /127.13)
Citrate	2.54 (d, α , β CH ₂ /48.17); 2.69 (d, α' , β' CH ₂ /48.17)
Creatine	3.04 (s, NCH ₃ /39.66); 3.94 (s, NCH ₂ /56.55)
Creatinine	3.05 (s, NCH ₃ /32.86); 4.06 (s, NCH ₂ /59.05)
Dimethylamine (DMA)	2.73 (s, CH ₃ /37.41)
Dimethylglycine (DMG)	2.93 (s, (CH ₃) ₂ /46.21); 3.72 (s, CH)
Ethanolamine	3.15 (t, CH ₂ NH ₂ /43.84), 3.84 (t, CH ₂ OH/60.01)
Formate	8.46 (s, CH/173.97)
Fumarate	6.53 (s, CH)
Galactose	3.49 (dd, C4H), 3.64 (dd, C3H), 3.75 (m, C1H, C2H, CH ₂), 3.83 (m, C3H), 3.93 (d, C2H), 3.98 (d, C2H), 4.10 (t, C1H), 4.60 (d, CH ₂), 5.28 (d, C5H)
α -Glucose	3.23 (dd, C2H); 3.44 (m, C4H); 3.50 (t, C3H); 3.72 (dd, C6H') ; 3.90 (m, C6H); 4.65 (d, C1H)
β -Glucose	3.42 (t, C4H); 3.54 (dd, CH); 3.71 (t, C3H); 3.77 (dd, C6H); 3.84 (m, C5H) ; 5.25 (d, C1H/94.97)
Glutamine	2.15 (m, β CH ₂ /29.24); 2.47 (m, γ CH ₂ /33.67); 3.79(t, α CH/57.41)
Glycine	3.57 (s, α CH ₂ /44.45)
Hippurate	3.97 (d, CH ₂); 7.56 (t, C4H, C6H ring); 7.64 (t, C3H, C5H ring); 7.83(d, C4H); 8.52 (br, NH)
Histidine	3.18 (dd, β CH ₂); 3.28 (dd, β' CH ₂); 4.01(dd, α CH ₂); 7.13 (s, C4H ring); 7.98 (s, C2H ring)
Hypoxanthine	8.20 (s, C2H ring); 8.22 (s, C8H)
Indoxyl sulphate (IS) *	7.21 (dd, C8H), 7.28 (dd, C7H), 7.36 (s, C2H), 7.51 (d, C6H), 7.70 (d, C9H)
Lactate	1.34 (d, CH ₃ /22.43); 4.11 (q, CH/71.53)
Lactose	3.28 (dd, C2H); 3.55 (m, C'2H); 3.59 (dd, C2H); 3.66 (m, C'3H, C3H, C5H); 3.73 (m, C'6H, C'5H); 3.79 (m, C6H); 3.86 (m, C6H, C3H); 3.94 (m, C6H, C'4H, C4H); 4.46 (d, C'1H/ 105.8); 5.25 (d, C1H)
Lysine	1.48 (m, γ CH ₂ /24.28); 1.73 (m, δ CH ₂ /29.06); 1.92 (m, β CH ₂ /32.60); 3.03 (t, ϵ CH ₂ /42.02); 3.77 (t, α CH/57.28)
Methylguanidine	2.83 (s, CH ₃ /29.68)
Myo-inositol	3.29 (t, C5H/76.87), 3.55 (dd, C1H, C3H/73.55), 3.63 (t, C4H, C6H/74.94), 4.08 (t, C2H/74.85)
<i>N</i> -methyl-2-pyridone-5-carboxamide (2-Py)	3.62 (s, CH ₃); 6.66 (d, C3H ring/120.85); 7.97 (dd, C4H ring); 8.33 (d, C6H ring/145.46)
<i>N</i> -methylnicotinamide (NMND)	4.48 (s, NCH ₃ /51.30); 8.18 (m, C5H ring); 8.90 (d, C4H ring); 8.97 (d, C6H ring); 9.29 (s, C2H ring)
Phenylacetylglutamine (PAG) *	1.93 (m, β CH ₂) ; 2.11 (m, β' CH ₂); 2.27 (t, γ CH ₂); 3.67 (d, CH ₂); 4.18 (m, α CH); 7.36 (m, C2H, C4H, C6H ring); 7.43 (m, C3H, C5H ring)
Pyruvate	2.39 (s, CH ₃)
Scyllo-inositol	3.36 (s, CH)
Succinate	2.41 (s, CH ₂ /36.85)
Taurine	3.26 (t, CH ₂ SO ₃); 3.43 (t, NCH ₂)
Threonine	1.33 (d, CH ₃ /22.44); 3.61 (d, β CH/63.36); 4.26 (dd, α CH/68.90)
Trigonelline	4.44 (s, CH ₃); 8.09 (t, C3H ring); 8.84 (br, C2H, C4H ring); 9.12 (s, C6H ring)

Assignment (<i>continues</i>)	δ_{H} ppm (multiplicity, assignment/ δ_{C} ppm)
Trimethylamine (TMA)	2.89 (s, CH ₃ /47.40)
Trimethylamine-N-oxide (TMAO)	3.27 (s, CH ₃ /62.31)
Tyrosine	3.06 (dd); 3.21 (dd); 3.95 (dd); 6.91 (d, C3H, C5H ring); 7.20 (d, C2H, C6H ring)
Urea	5.79 (br s, NH ₂)
Xylose	3.23 (dd, C3H), 3.33 (dd, C6H), 3.42 (t, C4H), 3.53 (dd, C3H), 3.63 (m, C6H, C5H, C4H), 3.93 (dd, C6H), 4.59 (d, C2H), 5.21 (d, C2H)

6.2 Impact of delivery mode and gender in the urine of healthy newborns

The goal of the work presented in this section was to determine the impact of possible confounding effects in newborn urine, i.e. to identify factors that could introduce variability in normal/healthy situations thus conditioning the results obtained for the subsequent study of disorders. In order to achieve this, control samples were used to compare the effect of a) delivery mode: vaginal delivery (VD) vs. cesarean section (CS), b) gender: male vs. female, c) day of life at sampling: day 1 vs. day 2 (no comparison was possible for subsequent days due to low number of samples available for other days of life), and d) gestational age (GA) at birth: 37 vs. 38 gestational weeks (g.w.), 38 vs. 39 g.w. and 39 vs. 40 g.w.. Sample numbers for each of the condition investigated are listed in Table 2.5 of Experimental Section.

The impact of each confounding effect was evaluated through PCA and PLS-DA, considering the original spectra and after variable selection using the method presented in Chapter 4. All PLS-DA models were submitted to validation through MCCV and evaluated in terms of robustness, determined through classification rates (CR), specificity and sensitivity, and predictive power, given by Q^2 values distribution of true and permuted models. The resulting model quality parameters and MCCV results are summarized in Table A-III.4, Annex III, the best models being consistently obtained by using the reduced dataset and noted in bold in the same table. Among the confounding effects investigated in this work, delivery mode was found to cause the highest impact in newborn urine, followed by gender, both models being subjected to further analysis and presented in detail below. Regarding day of life at sampling, specifically day 1 vs. day 2, the PLS-DA model was found to have relatively low robustness and predictive power given by poor ROC space result and Q^2 values distribution (not shown). Previous urine metabolomic studies of healthy term newborns had shown significant decreases in NMND, taurine and bile acids

and increases in betaine from day 1 to day 6 (Trump *et al.*, 2006), but did not discuss changes specifically from day 1 to day 2. Nevertheless, the reduced number of samples available for day 1 ($n=7$) compared to day 2 ($n=31$) can be compromising the result obtained in this work and, hence, higher number of samples are necessary, preferably including a larger range of days, so that more definitive conclusions can be drawn.

6.2.1 Effects of delivery mode on urine composition: results and proposed metabolic interpretation

In spite of the acknowledgement of the importance of the birth process on future health, there is scarce literature focused on the impact of delivery mode in the metabolic adaptations to post-natal life. A few reviews have reported that CS seems to be associated with higher respiratory morbidity, namely the need for respiratory support due to impaired lung adaptation mediated during labor (Boutsikou and Malamitsi-Puchner, 2011), and significant changes in gut microflora which can be related to later decrease immune responses (Boutsikou and Malamitsi-Puchner, 2011, Hyde *et al.*, 2012), development of allergic rhinitis and asthma (Funkhouser and Bordenstein, 2013), increased risk of childhood onset Type 1 Diabetes (Cardwell *et al.*, 2008, Boutsikou and Malamitsi-Puchner, 2011), celiac disease and inflammatory bowel disease (Funkhouser and Bordenstein, 2013).

Results of delivery mode effects

In this work, urine from newborns born by VD and CS were clearly separated in the PCA and PLS-DA score plots, as shown in Figure 6.2a,b, the latter model having satisfactory ROC results (CR of 88%, sensitivity of 79% and specificity 94%) along with low overlapping of Q^2 values of true and permuted models, as shown in Figure 6.2c,d. Regarding the CS sample noted in Figure 6.2b (see arrow), inspection of the corresponding NMR spectrum revealed high resonance of 4-hydroxyphenylacetate (4-HPA), phenylacetylglutamine (PAG) (both metabolites being increased in the VD group) along with high mannose and an unassigned broad resonance at 5.52 ppm. The metabolic profile associated with delivery mode was given by the subset of variables used in the PLS-DA model, this profile being registered in Table 6.2. Delivery mode caused changes in 15

metabolites, mainly comprising metabolites derived from the gut microflora and organic acids, 3 unassigned singlets (at 0.56, 2.06 and 3.96 ppm) and 9 unassigned spectral regions (i.e. regions with unidentifiable and/or broadly overlapped resonances) with $p < 0.05$. At the end of this table (and as will be done for other urinary profile tables presented in this chapter) the relevant literature review of delivery mode related changes is shown for comparison.

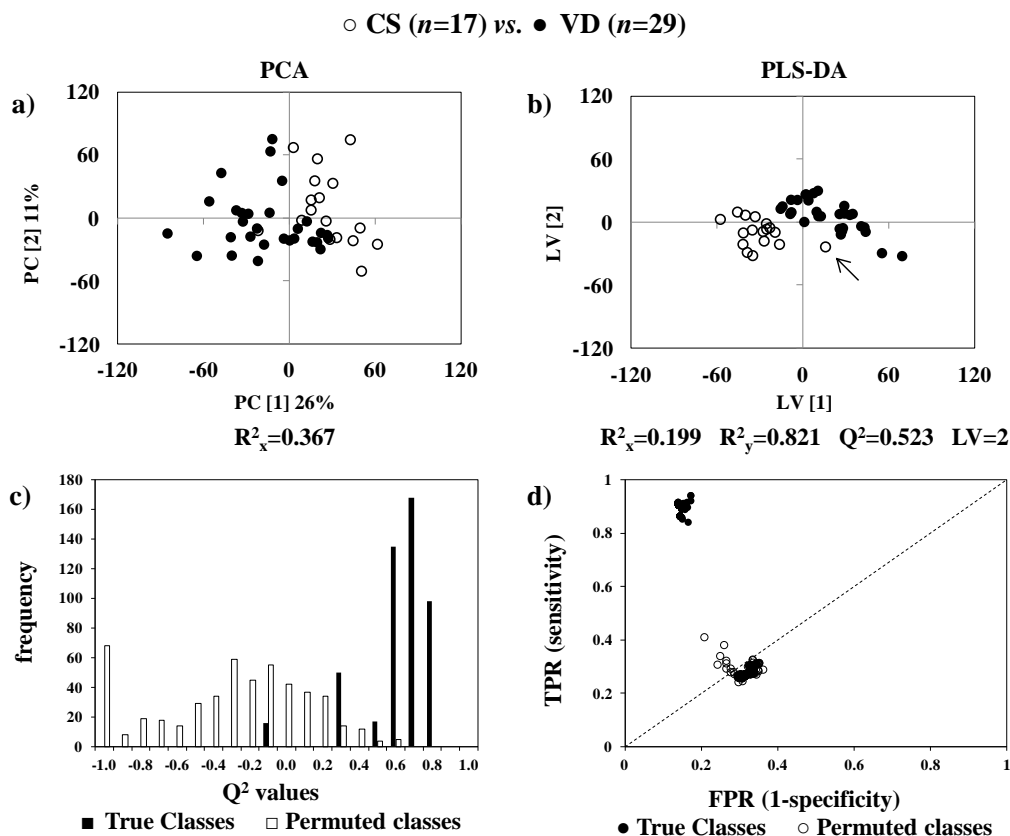


Figure 6.2: a) PCA and b) PLS-DA score plots, c) Q^2 values and d) ROC plot of true and permuted models obtained for urine of newborns born through cesarean section (CS, $n=17$) compared to vaginal delivery (VD, $n=29$). TPR: true positive rate, FPR: false negative rate. Arrows notes outlier sample.

To facilitate visual evaluation of changes related to delivery mode, the urinary profile was plotted in terms of each variable's VIP by means of a VIP-wheel representation (Figure 6.3a). Moreover, each integral was represented in a Volcano plot, shown in Figure 6.3b. The Volcano plot can be regarded as a graphical summary of the biological and statistical significance given respectively by the effect size and the p -value (Cui and Churchill, 2003). One advantage of using this plot is the easiness of interpretation, as it is easy to visually identify the most relevant metabolites/resonances which are located in the

Table 6.2: Metabolite/resonance changes in newborn urine of newborns born through vaginal delivery (VD) compared to cesarean section (CS). ^a Chemical shifts selected by variables selection; s: singlet, d: doublet, t: triplet, m: multiplet. ^b Peaks possibly related to diet and/or gut microflora. Ui: unassigned compound *i* by order of appearance. 2-KG: 2-ketoglutarate, 4-DEA: 4-deoxyerythronic acid, 4-HPA: 4-hydroxyphenylacetate, DMA: dimethylamine, IS: indoxyl sulphate, PAG: phenylacetylglutamine. Only *p*-values<0.05 are shown.

Impact of delivery mode in the newborn urine profile		
CS (<i>n</i> =17) vs. VD (<i>n</i> =29)		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
1-Me-histidine ^b	7.04 s, 7.78 s	↑(0.50)
2-KG	2.45 t	↑(0.09)
4-DEA	1.11 d	↓(-0.14)
4-HPA ^b	7.17 d	↓(-0.56)
Acetone	2.23 s	↑(0.68, 1.64×10 ⁻²)
Betaine	3.26 s, 3.91 s	↑(0.51)
Cis-aconitate	3.12 d	↑(0.44)
DMA ^b	2.72 s	↑(0.61, 4.64×10 ⁻²)
Ethanolamine	3.84 t	↑(0.47)
IS ^b	7.51 d, 7.70 d	↓(-1.01, 1.28×10 ⁻³)
Lactose	4.46 d, 5.25 d, 3.28 t, 3.55 m, 3.59 d, 3.66 m, 3.73 t, 3.79 m, 3.94 m	↑(0.61)
Myo-inositol	3.55 dd	↓(0.29)
PAG ^b	7.36 m, 7.43 m	↓(-0.71, 1.35×10 ⁻²)
Trigonelline ^b	9.12 s, 8.84 m	↓(-1.34, 2.35×10 ⁻²)
Tyrosine	6.91 d	↓(-0.44)
U1	0.56 s	↓(-0.71, 2.98×10 ⁻²)
U2	2.06 s	↑(2.04, 2.38×10 ⁻³)
U3	3.96 s	↑(0.93, 1.64×10 ⁻²)
Unassigned spectral regions		
0.87-0.89		↓(-0.61, 3.73×10 ⁻²)
2.90-2.92		↑(0.80, 9.17×10 ⁻³)
2.94-2.99		↑(1.05, 1.18×10 ⁻³)
3.19-3.20		↑(0.97, 3.71×10 ⁻³)
3.22-3.24		↑(0.5, 4.71×10 ⁻²)
3.97-4.01		↑(0.75, 2.98×10 ⁻²)
6.71-6.74		↓(-0.71, 3.73×10 ⁻²)
7.24-7.30		↓(-0.91, 3.38×10 ⁻⁴)
8.64-8.67		↓(-0.80, 1.54×10 ⁻³)
+ 4 Unassigned spin and 5 Regions with <i>p</i> >0.05		
Literature review:		
CS without labor compared to VD: ↓ isoleucine, fructose, mannose, glucose, allose, glucuronic acid, inositol and cysteine in umbilical cord blood (Hashimoto <i>et al.</i> , 2013); CS compared to VD: ↑ hepatic glycerol, ↓ glycerol phosphate dehydrogenase activity (in piglets liver) (Hyde <i>et al.</i> , 2010)		

upper lateral side of the plot. This means that metabolites positioned in the upper right or (if increased or decreased, respectively) part of the plot have simultaneously higher biological importance (i.e. higher effect size), and higher statistical significance (given by

the $-\log_{10}$ of the p -value). In the specific Volcano plot shown in Figure 6.3b, metabolites/resonances on the right side of the plot are increased in newborns born through CS whilst metabolites in the left side of the plot are increased in VD.

Visual inspection of the VIP-wheel and Volcano plot (Figure 6.3a,b) of the effect of delivery mode showed that newborns born through CS had higher contents of acetone and DMA along with an unassigned singlet at 2.06 and 3.96 ppm and unassigned spectral regions 2.94-2.99, 3.19-3.20, 2.90-2.92, 3.97-4.01, 0.87-0.89 and 3.22-3.24 ppm (aliphatic and bile acids regions). On the other hand, resonances of IS, PAG and trigonelline, metabolites known to derive from gut microflora, were found to be highly relevant for the VD profile along with higher contents of the unassigned singlet at 0.56 ppm and spectral regions 7.24-7.30, 8.64-8.67, 6.71-6.74 ppm (aromatic regions). Peaks arising from metabolites deriving from the gut microflora are typically located in the aromatic region (6-9 ppm) of the urine spectra, this region indeed contributing significantly for the separation obtained with the PLS-DA model, as shown in the VIP-wheel and Volcano plot in Figure 6.3a,b.

Proposed metabolic interpretation of changes related with delivery mode

Differences found, in this study, in newborn urine according to delivery mode are consistent with previous knowledge on gut microflora colonization in VD vs. CS, increased stress associated with labor and possibly altered hepatic metabolism. Delivery mode plays an important role in the selecting the first the gut microflora colonizers, with babies born through CS having reduced number of bacteria and delayed the appearance of some bacterial species (Morelli, 2008). This different gut microflora colonization in babies born through VD can in fact explain the increase in gut derived metabolites PAG and trigonelline here found. Another interesting finding regards the significant increase in urinary excretion of IS in VD newborns. IS is formed in the liver from indole, a product of intestinal tryptophan breakdown and sometimes related with microflora metabolism (Suhre *et al.*, 2010, Heinzmann *et al.*, 2012). Besides the possible microflora connection, IS has also been found to mediate oxidative stress *in vitro* in endothelial cells (Dou *et al.*, 2007) and to induce nephrotoxicity (Suhre *et al.*, 2010). With regard to delivery mode, increased oxidative stress has indeed been suggested in neonates born through VD compared to CS

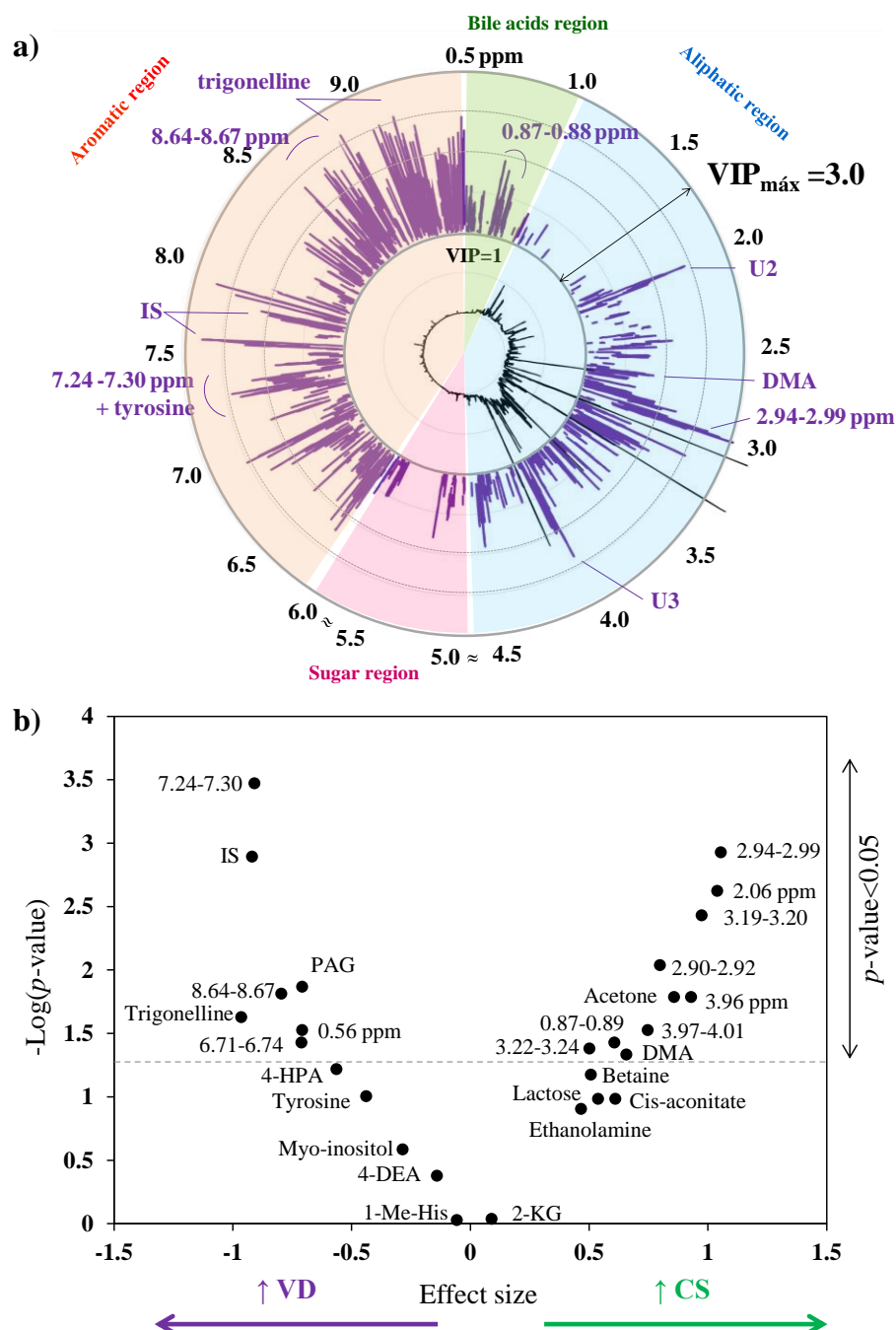


Figure 6.3: a) VIP-wheel representation of the NMR urinary signature of newborns born through vaginal delivery vs. cesarean section (each point in the ppm scale represents a selected variable with corresponding VIP value in the radial scale). The inner and outer circles represent, respectively, the average urine spectrum of healthy newborn urine (control group) and corresponding ppm scale. Main differences between VD and CS are noted. b) Volcano plot of effect size vs. $-\text{Log}(p\text{-value})$ of metabolite/resonances integrals listed in Table 6.2. Dashed line indicates $p\text{-value}=0.05$.

(Vakilian *et al.*, 2009, Hashimoto *et al.*, 2013, Watanabe *et al.*, 2013). Thus, higher urinary IS here found in VD can be a reflection of delivery mode related stress, although a concomitant contribution of gut microflora metabolism needs to be acknowledged.

Regarding the CS group, a significant increase was found for acetone, U2 and U3 (singlets at 2.06 and 3.96 ppm) and unassigned resonances in the aliphatic region (mainly at 2.90-3.20 ppm). In relation to acetone, ketone bodies are formed in the liver during fatty acids β -oxidation producing acetyl-CoA that enters the TCA cycle. A previous animal study showed that piglets born through CS had profound changes in liver metabolism, and showed evidence of decreased gluconeogenesis from glycerol (Hyde *et al.*, 2010). Although no concomitant changes were found for other ketone bodies, the acetone increase seen here can reflect changes in hepatic fatty acids β -oxidation and gluconeogenesis, consistently with changes previously found in the animal model.

6.2.2 Effect of gender on urine composition: results and proposed metabolic interpretation

Previous metabolomic studies have unveiled gender related differences in adults' urine these being associated with differences in muscle mass, leading to altered excretion of creatine and creatinine (Slupsky *et al.*, 2007, Psihogios *et al.*, 2008), and hormonal effects, causing differences in mitochondrial metabolism specifically in TCA cycle (Kochhar *et al.*, 2006, Slupsky *et al.*, 2007), methylamines (Kochhar *et al.*, 2006, Psihogios *et al.*, 2008) and lipids metabolism, specifically in β -oxidation of fatty acids (Slupsky *et al.*, 2007). The list of metabolites reported in literature to change in relation to gender is listed in Table 6.3 below.

Results of gender effects

Gender-related differences were found to pose the second highest impact in newborn urine composition, following delivery mode. A clear separation between urine of male and female newborns was obtained through PCA and further confirmed through PLS-DA as shown respectively in Figure 6.4a,b. The PLS-DA model was found to be robust and have predictive power, determined by the low overlap of Q^2 values, CR 76%, sensitivity 75% and specificity 76%, shown in Figure 6.4c, d.

Table 6.3: Urinary changes in relation to gender in adult cohorts: a literature review.

Compound and variation		Reference
3-HBA	↑Male	(Psihogios <i>et al.</i> , 2008)
Acetone	↑Male	(Slupsky <i>et al.</i> , 2007)
Acetylcarnitine	↑Male	(Slupsky <i>et al.</i> , 2007)
Carnitine	↑Male	(Slupsky <i>et al.</i> , 2007)
Citrate	↑Female	(Kochhar <i>et al.</i> , 2006, Slupsky <i>et al.</i> , 2007, Psihogios <i>et al.</i> , 2008)
Creatine	↑Female	(Slupsky <i>et al.</i> , 2007, Psihogios <i>et al.</i> , 2008)
Creatinine	↑Male	(Slupsky <i>et al.</i> , 2007, Psihogios <i>et al.</i> , 2008)
DMA	↑Male	(Kochhar <i>et al.</i> , 2006, Psihogios <i>et al.</i> , 2008)
	↑Female	
Fumarate	↑Female	(Slupsky <i>et al.</i> , 2007)
Glutamic acid	↑Male	(Ramautar <i>et al.</i> , 2011)
Glycine	↑Female	(Psihogios <i>et al.</i> , 2008)
Hippurate	↑Female	(Psihogios <i>et al.</i> , 2008)
Hypotaurine	↑Male	(Ramautar <i>et al.</i> , 2011)
Lactate	↑Female	(Psihogios <i>et al.</i> , 2008)
Methionine	↑Male	(Ramautar <i>et al.</i> , 2011)
Methylhistidine	↑Male	(Ramautar <i>et al.</i> , 2011)
NMND	↑Male	(Ramautar <i>et al.</i> , 2011)
Phenylalanine	↑Female	(Psihogios <i>et al.</i> , 2008)
Proline betaine	↑Male	(Ramautar <i>et al.</i> , 2011)
Pyroglutamic acid	↑Male	(Ramautar <i>et al.</i> , 2011)
Succinate	↑Female	(Psihogios <i>et al.</i> , 2008)
Taurine	↑Male	(Kochhar <i>et al.</i> , 2006, Psihogios <i>et al.</i> , 2008)
Threonine	↑Male	(Ramautar <i>et al.</i> , 2011)
TMAO	↑Male	(Psihogios <i>et al.</i> , 2008)

The analysis of the subset of variables used in the PLS-DA model enabled a gender-specific metabolic signature to be unveiled, this being registered in Table 6.4. Changes were found for 20 identified metabolites, including 3 sugars, 3 amino and 4 organic acids, 1 ketone body, 3 products of nucleotide metabolism, 4 gut microflora derived metabolites, 2 others (creatine and *myo*-inositol), 6 unassigned spin systems (triplet at 5.35 ppm, doublet at 5.41 ppm, multiplet at 5.45 ppm, doublet at 5.59 ppm, doublet at 6.07 ppm and singlet at 7.93 ppm) and 11 spectral regions with $p < 0.05$. Out of these variations, 3 metabolites (namely acetone, DMA and fumarate) were found consistently with previous gender-specific changes reported in adults, these being noted with ^c in Table 6.4. On the other hand, 5 metabolites were found to vary in opposite direction to that reported in adult cohorts (indicated with ^d in the same table), namely decreased creatine, hippurate, and increased NMND, taurine and threonine found increased in this study in females the opposite direction being reported in female adults.

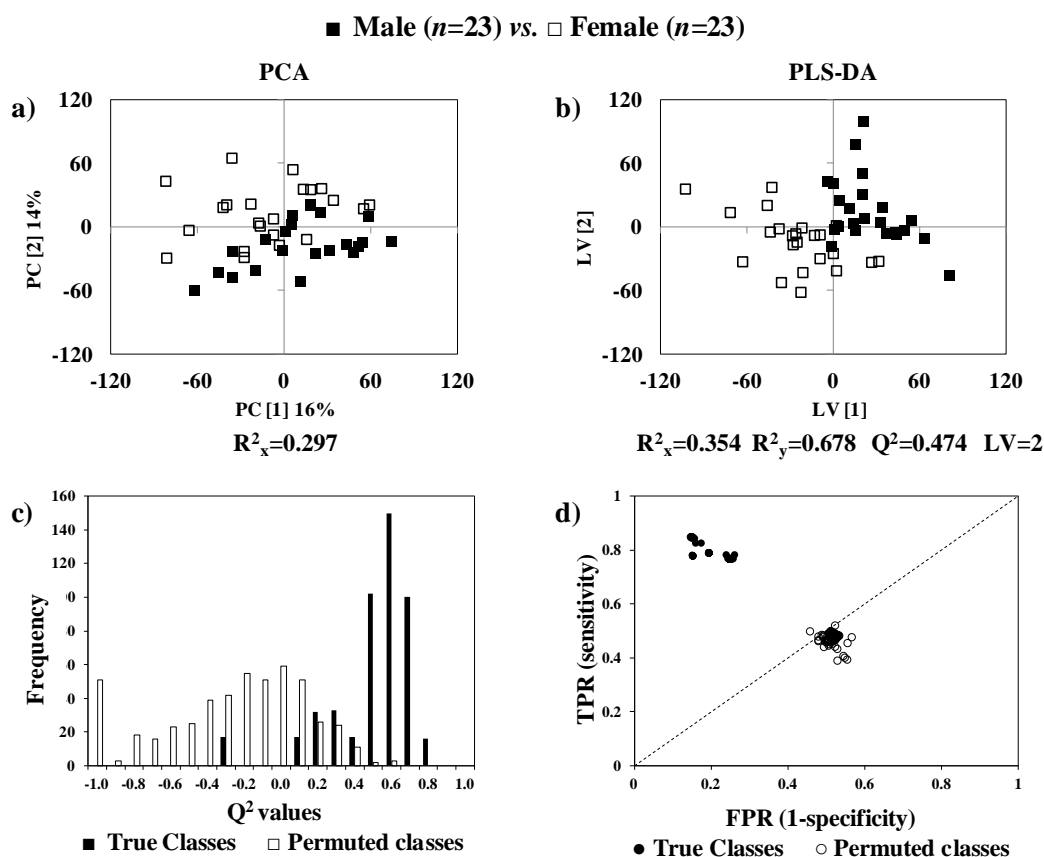


Figure 6.4: a) PCA and b) PLS-DA score plot, c) Q^2 values and d) ROC plot of true and permuted models obtained for newborn urine of males ($n=23$) compared to females ($n=23$). TPR: true positive rate, FPR: false negative rate.

To enable visual identification of the most relevant metabolites/regions varying between male and female newborns, the urinary profile was plotted in a VIP-wheel (Figure 6.5a), which shows that metabolites with higher importance (VIP) are *myo*-inositol, lysine, taurine, allantoin, U9 (putatively xanthine), resonances at the sugar region and unassigned region 6.11-6.48 ppm (typically containing resonances of purines and pyrimidines (Wevers *et al.*, 1999)). Also, the effect size and p -value for each metabolite/region were plotted in a Volcano plot (Figure 6.5b) highlighting resonances with simultaneously high biological (effect size) and statistical (p -value) significance. In this figure, metabolites increased in females are located in the right hand side of the plot whilst metabolites increased in males are positioned in the left hand side of the plot.

Table 6.4: Metabolite/resonance changes in newborn urine of males compared to females. ^a Chemical shifts selected by variables selection; TOCSY or STOCYSY correlated peaks in square brackets; s: singlet, d: doublet, t: triplet, m: multiplet, br: broad. ^b Peaks possibly related to diet and/or gut microflora. ^c metabolites varying consistently in nature and direction and ^d in opposite direction to gender-related changes reported in adults (Kochhar *et al.*, 2006, Slupsky *et al.*, 2007, Psihogios *et al.*, 2008, Ramautar *et al.*, 2011). Ui: unassigned compound i by appearance order and following Table 6.2. NMND: *N*-methylnicotinamide, other abbreviations defined in Table 6.2. *Putatively assigned as xanthine. Only *p*-values<0.05 are shown.

Impact of gender in the newborn urine profile		
Female (<i>n</i> =23) vs. Male (<i>n</i> =23)		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
2-KG	2.45 t, 3.01 t	↑(0.29)
4-OH-hippurate ^b	6.98d	↓(-0.24)
Acetone ^c	2.23 s	↓(-0.58)
Allantoin	5.39 s	↑(0.81, 1.3×10 ⁻³)
<i>Cis</i> -aconitate	3.12 d	↑(0.26)
Creatine ^d	3.04 s	↓(-0.39)
DMA ^{b, c}	2.72 s	↓(-0.36)
Formate	8.46 s	↓(-0.45)
Fumarate ^c	6.53 s	↑(0.34)
Galactose	4.59 d, 5.28 d	↓(-0.15)
Glucose	3.24 dd, 3.92 dd, 5.25 d	↑(0.83, 1.2×10 ⁻²)
Hippurate ^{b, d}	7.56 t, 7.64 t, 7.84 d	↓(-0.40)
Lactose	5.25 d, 3.55 m, 3.59 d, 3.66 m, 3.73 t,	↑(0.93, 3.6×10 ⁻²)
Lysine	1.73 m	↓(-0.71, 3.6×10 ⁻²)
Methylguanidine	2.83 s	↑(0.49)
<i>Myo</i> -inositol	3.29 t, 3.55 dd, 3.63 t, 4.08 t	↓(-0.86, 4.9×10 ⁻³)
NMND ^d	4.48 s, 8.18 m, 8.90 d, 8.98 d, 9.28 s	↑(0.24)
<i>Scyllo</i> -inositol	3.36 s	↓(-0.61, 1.8×10 ⁻²)
Taurine ^d	3.43 t	↑(0.69, 4.5×10 ⁻²)
Threonine ^d	4.26 dd	↑(0.58)
U4	5.35 t [3.64]	↑(0.67, 2.7×10 ⁻²)
U5	5.41 d [3.56, 3.78]	↑(0.72, 6.3×10 ⁻⁶)
U6	5.45 m [2.84]	↓(-0.72, 1.1×10 ⁻²)
U7	5.59 d [1.61 m, 2.06, 2.96]	↓(-0.75, 1.8×10 ⁻²)
U8	6.07 d	↓(-0.71, 1.7×10 ⁻²)
U9*	7.93 s	↑(1.01, 3.1×10 ⁻³)
Unassigned spectral regions		
1.57-1.66		↓(-0.57, 4.0×10 ⁻²)
2.17-2.20		↓(-0.65, 4.5×10 ⁻²)
2.80-2.82		↑(0.54, 8.6×10 ⁻³)
4.32-4.34		↑(0.74, 1.6×10 ⁻²)
6.11-6.12		↓(-0.85, 1.3×10 ⁻²)
6.30-6.34		↓(-0.75, 2.0×10 ⁻²)
6.36-6.38		↓(-0.63, 1.8×10 ⁻²)
6.39-6.41		↓(-0.84, 1.1×10 ⁻²)
6.46-6.49		↓(-1.09 1.5×10 ⁻³)
6.60-6.62		↓(-0.83, 1.3×10 ⁻²)
8.03-8.07		↑(0.81, 1.2×10 ⁻²)
+ 6 Unassigned spin and 6 regions with <i>p</i> >0.05		

Literature review in urine of adults:

↑Male: 3-HBA (Psihogios *et al.*, 2008), acetone (Slupsky *et al.*, 2007), creatinine (Slupsky *et al.*, 2007, Psihogios *et al.*, 2008), DMA (Kochhar *et al.*, 2006, Psihogios *et al.*, 2008), NMND (Ramautar *et al.*, 2011), taurine (Kochhar *et al.*, 2006, Psihogios *et al.*, 2008), threonine (Ramautar *et al.*, 2011), TMAO (Psihogios *et al.*, 2008); ↑Female: citrate (Kochhar *et al.*, 2006, Slupsky *et al.*, 2007, Psihogios *et al.*, 2008), creatine (Slupsky *et al.*, 2007, Psihogios *et al.*, 2008), fumarate (Slupsky *et al.*, 2007), glycine (Psihogios *et al.*, 2008), hippurate (Psihogios *et al.*, 2008), lactate (Psihogios *et al.*, 2008), succinate (Psihogios *et al.*, 2008).

Inspection of the VIP-wheel and Volcano plot (Figure 6.5) showed that female newborns have higher contents of an unassigned doublet at 5.41 ppm, allantoin, glucose, lactose, taurine, U9 (putatively xanthine) and other unassigned spin systems and spectral regions. On the other hand, male newborns were found to have higher *myo*- and *scyllo*-inositol and lysine along with several unassigned regions and spin systems from 5.45 to 6.62 ppm.

Proposed metabolic interpretation of changes related with gender

In this study, variations in DMA, fumarate, and acetone were found commonly (in nature and direction) with previously reported variations in adult cohorts whilst creatine, hippurate, NMND, taurine and threonine varied in opposite direction (Kochhar *et al.*, 2006, Slupsky *et al.*, 2007, Psihogios *et al.*, 2008, Ramautar *et al.*, 2011). Additionally, other 16 metabolites were reported in literature to be gender dependent, however, only 8 of those were here identified in newborn urine but not found to vary in this cohort. These findings suggest a different gender-related effect in adults and in newborns.

In this cohort, metabolites/resonances with higher importance for class separation as well as statistical and biological relevance are shown in Figure 6.5a,b. Urine from females was characterized by higher allantoin, unassigned doublet at 5.41 ppm (U5) and unassigned singlet at 7.93 ppm (U9). Through comparison with standards databases (Bruker biorefcode and HMDB (Wishart *et al.*, 2009) the singlet at 7.93 ppm was found to possibly belong to xanthine. Still, due to the low intensity of this resonance, definitive confirmation with HSQC was not possible at this stage thus remaining as putative assignment. Allantoin is an end product of purine metabolism, formed from oxidation of xanthine to uric acid (Caussé *et al.*, 2007), as shown in Figure 6.6. The significant and concomitant increases of allantoin and putatively for xanthine suggest altered purine degradation in females, although no metabolic explanation corroborating this hypothesis was found in literature.

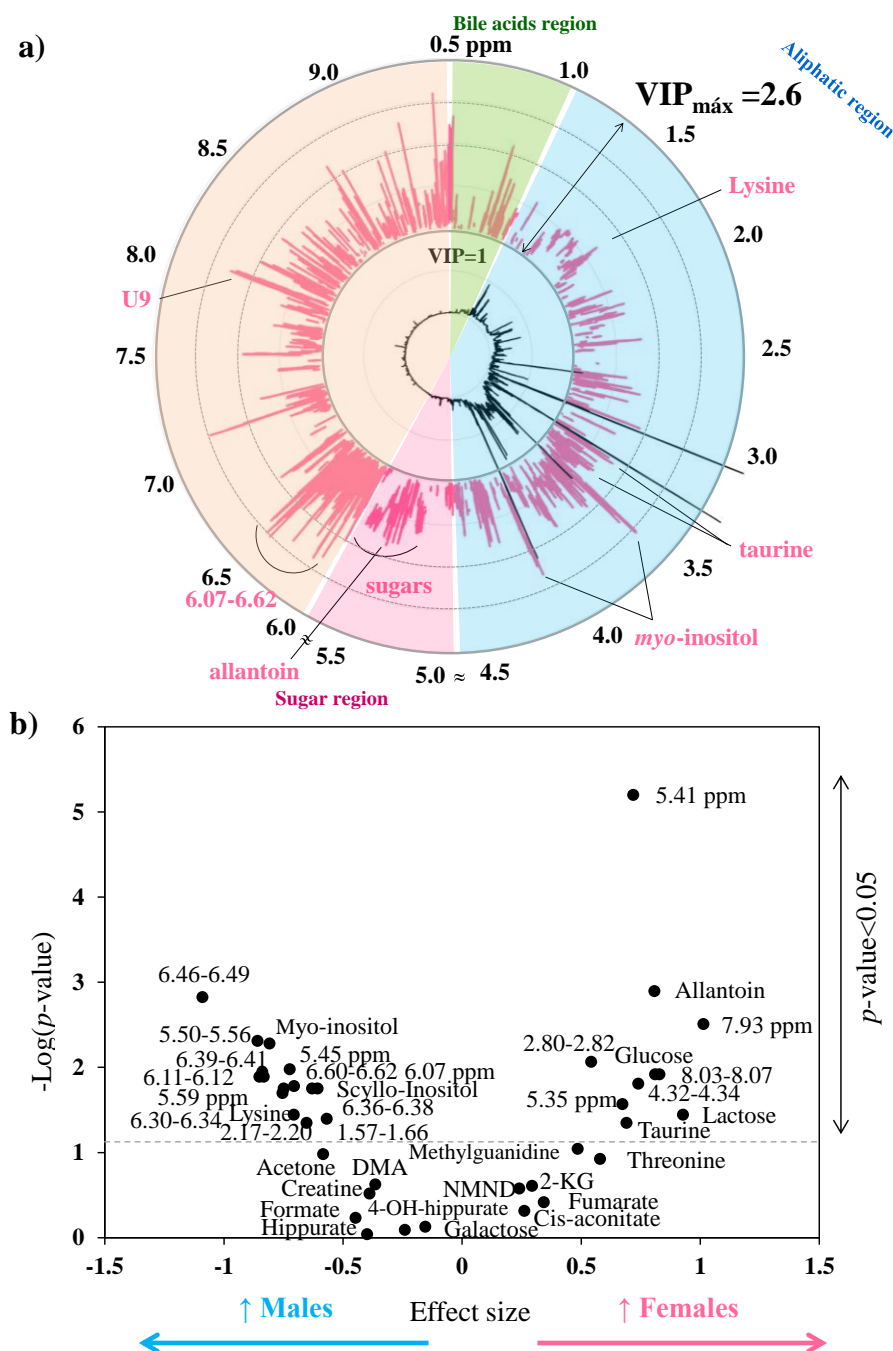


Figure 6.5: a) VIP-wheel representation of the NMR urinary signature M vs. F (each dot represents a selected variable with corresponding VIP value in the radial scale). The inner and outer circles represent, respectively, the average urine spectrum of healthy newborn urine (control group) and ppm scale. Main differences between male and female are noted. b) Volcano plot of effect size vs. $-\text{Log}(p\text{-value})$ of metabolite/resonances integrals' listed in Table 6.4. Dashed line indicates $p\text{-value}=0.05$.

Regarding male newborns, significantly higher *myo*-inositol and metabolites resonating in spectral regions 5.50-5.56 ppm (sugars region) and 6.46-6.48 ppm (purines

and pyrimidines region (Wevers *et al.*, 1999)) were registered, the latter variation supporting the previous hypothesis of altered purines degradation due to gender. *Myo*-inositol can be endogenously synthesized from glucose (Croze and Soulage, 2013) (glucose being found to decrease in females) and is converted to *scyllo*-inositol (also found increased in females) in the gut microflora. These findings suggest an apparent alteration in sugars metabolism associated with gender in the first days of life. As no literature was found supporting differences in gender-related sugar and nucleotide metabolism in adults, it can be suggested that these differences are potentially unique to this early age.

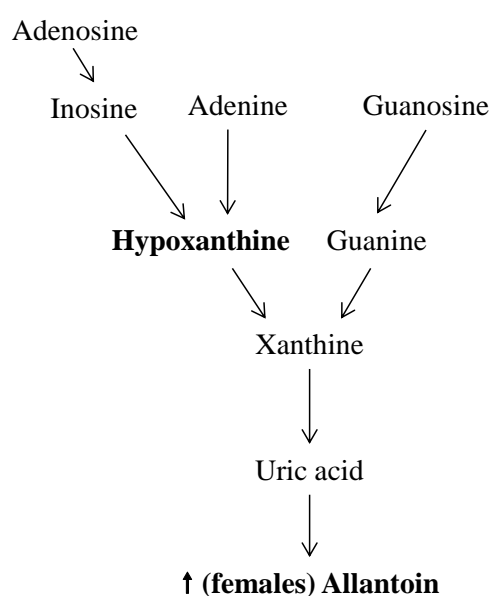


Figure 6.6: Degradation pathways of purines. Metabolites detected in urine are noted in bold. The small arrow indicate gender-related variation found in this study.

6.3 Impact of newborns disorders on newborn urine composition

6.3.1 Prematurity

Preterm birth is the major cause of perinatal morbidity and mortality (Hacker *et al.*, 2004) with 24% of neonatal deaths in developed countries being related to complications of prematurity (Kliegman *et al.*, 2011). Common neonatal complications due to prematurity include respiratory problems (respiratory distress syndrome (RDS) and apnea), cardiovascular (patent ductus arteriosus and bradycardia), gastrointestinal (poor gastrointestinal function and hyperbilirubinemia), metabolic/endocrine (hypocalcemia,

hypoglycemia, hyperglycemia and hypothermia), central nervous system complications (intraventricular hemorrhage and hypotonia), renal complications (hyponatremia, hypernatremia, hypokalemia) and infection risk (Kliegman *et al.*, 2011). The assessment of long-term effect of being born prematurely has inspired a few studies, these having shown alterations in brain metabolites and structure in adolescents (Gimenez *et al.*, 2008) and lipids metabolism in young adults born premature (Thomas *et al.*, 2011).

6.3.1.1 Unveiling an urinary metabolic signature of premature newborns

In this study, the urine from premature newborns (born before the 37th g.w.) was compared to controls (the detailed description of samples metadata being show in Table 2.5 of Experimental Section). Visual comparison of the average ¹H NMR spectra of cases and controls (Figure 6.1a,b) enabled the identification of significant differences, namely for 3-HBA 3-aminoisobutyrate, acetone, *myo*-inositol, fumarate, formate and NMND, these being noted in the same figure. In order to confirm these changes and unveil a more complete metabolic picture of premature newborns, PCA and PLS-DA were applied to the original and reduced dataset after variable selection. PCA revealed a separation trend between groups (along PC1) this being confirmed through PLS-DA, as shown in the corresponding score plots in Figure 6.7a,b. One note regarding the outlier visible in the PCA score plot, this dot corresponding to a premature newborn and the sample being collected on the 6th day of life, born from a mother affected by Crohn's disease and hypertension during the last week of gestation. Visual inspection of the corresponding NMR spectrum revealed a higher concentration of acetate and 2-ketoglutarate (2-KG) in this sample, suggesting a perturbation of the TCA cycle and decreased NMND and higher resonances in the bile acids region, consistent with previous reports (Trump *et al.*, 2006), whilst revealing no changes in taurine and betaine as previously suggested from day 1 to day 6 (Trump *et al.*, 2006). A distinction was also noted for a control sample located within the premature samples in the PCA score plot (Figure 6.7a), this sample corresponding to the same outlier identified in the PLS-DA score plot of VD vs. CS (Figure 6.2). MCCV confirmed the PLS-DA model's predictive power, given by low overlapping of Q^2 of true and permuted models and median Q^2 0.603, and robustness,

supported by ROC plot, CR, sensitivity and specificity of 92%, as shown in Figure 6.7c,d and Table A-III.4, Annex III.

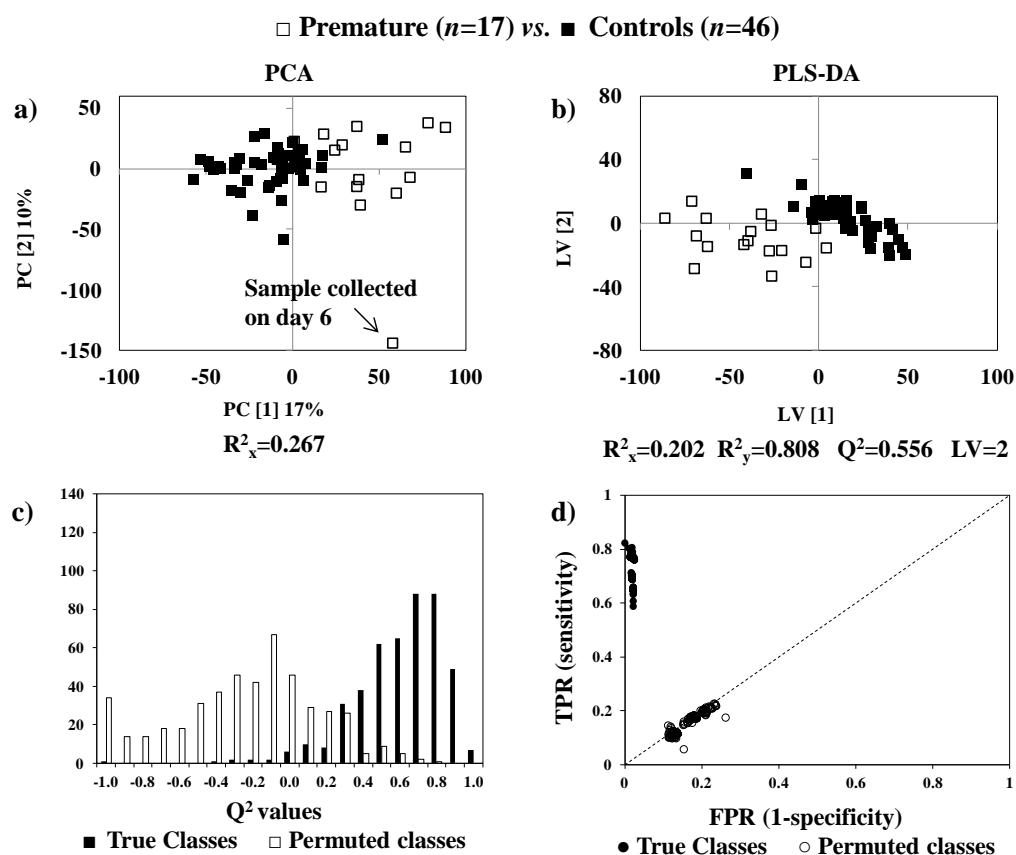


Figure 6.7: a) PCA and b) PLS-DA score plot, c) Q^2 values and d) ROC plot of true and permuted models obtained for urine of premature newborns ($n=17$) compared to controls ($n=46$). TPR: true positive rate, FPR: false negative rate.

The metabolic changes underlying the separation seen in the PLS-DA between premature newborns and controls, was further retrieved and analyzed in terms of biological and statistical significance (given by effect size and p -value, respectively). The overall changes unveiled in this study are listed in Table 6.5 with the corresponding variation (along with a summary of changes reported in literature in the last row for comparison). Relevant changes were found for 24 metabolites (comprising 5 organic acids, 3 products of nucleotide metabolism, 2 ketone bodies, 2 amino acids and derivatives, 2 metabolites involved in lung surfactant production, 3 markers of renal function, 5 metabolites derived from gut microflora and 2 others), 4 unassigned spin systems (singlets at 4.51, 6.59 and 7.93 ppm, the latter possibly belonging to xanthine, and a doublet at 6.76 ppm) and 5 unassigned spectral regions (with $p < 0.05$).

Table 6.5: Metabolite/resonance changes in newborn urine of premature newborns. ^a Chemical shifts selected by variables selection; TOCSY or STOCYSY correlated peaks in square brackets; s: singlet, d: doublet, t: triplet, m: multiplet. ^b Peaks possibly related to diet and/or gut microflora. ^c Consistent with previous reports (Foxall *et al.*, 1995, Constantinou *et al.*, 2005, Trump *et al.*, 2006), ^d Consistent with findings in 2nd T maternal urine of *pre*-PTD women (Section 4.2.1). Ui: unassigned compound i by order of appearance and following Table 6.2 and Table 6.4. 2-Py: *N*-methyl-2-pyridone-5-carboxamide, 3-HBA: 3-hydroxybutyrate, 3-HIVA: 3-hydroxyisovalerate, 4-DTA: 4 deoxythreonic acid, other abbreviations defined in Table 6.2 and Table 6.4. Only *p*-values<0.05 are shown.

Premature newborns (n=17) vs. Control (n=46)		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
1-Me-histidine ^b	7.78 s	↓(-0.83, 4.2×10 ⁻³)
2-Py	6.66 d, 8.33 d	↓(-0.96, 1.3×10 ⁻³)
3-aminoisobutyrate	1.19 d, 2.61 m, 3.06 dd	↓(-1.37, 7.5×10 ⁻⁷)
3-HBA	1.20 d	↓(-0.49, 4.8×10 ⁻³)
3-HIVA	1.27 s	↑(0.54, 4.3×10 ⁻²)
3-Me-histidine ^{b, d}	7.15 s, 8.12 s	↓(-0.53)
4-DTA	1.23 d	↓(-0.61, 4.8×10 ⁻²)
4-OH-hippurate ^{b, d}	7.76 d	↑(0.32)
Acetone	2.23 s	↓(-0.55)
<i>Cis</i> -aconitate	3.12 d	↓(-1.37, 2.6×10 ⁻⁶)
Citrate ^d	2.54 d, 2.69 d	↑ (0.45, 4.6×10 ⁻²)
Creatine	3.04 s, 3.93 s	↓(-0.58, 2.9×10 ⁻²)
Creatinine ^d	3.05 s, 4.06 s	↓(-0.85, 1.2×10 ⁻³)
DMA ^b	2.72 s	↓(-0.88, 5.5×10 ⁻³)
Ethanolamine	3.15 t, 3.84 t	↓(-0.97, 6.2×10 ⁻⁴)
Formate	8.46 s	↑(0.89, 5.2×10 ⁻³)
Fumarate	6.53 s	↑(0.42)
IS ^b	7.51 d	↑(1.00, 4.8×10 ⁻²)
Lactate	4.11 q	↓(-0.76, 6.7×10 ⁻³)
Lysine	1.73 m	↑(0.63, 2.4×10 ⁻²)
Methylguanidine	2.83 s	↓(-0.78, 4.7×10 ⁻³)
<i>Myo</i> -inositol ^c	3.26 t, 3.53 dd, 4.07 t	↑(0.87, 2.0×10 ⁻²)
NMND ^c	4.48 s, 8.18 m, 8.90 d, 8.97 d,	↑(1.20, 4.5×10 ⁻⁴)
Succinate	2.41 s	↑(0.61, 5.0×10 ⁻²)
U10	4.51 s	↑(0.67, 3.2×10 ⁻³)
U11	6.59 s	↑(1.48, 1.8×10 ⁻⁵)
U12	6.76 d	↓(-1.17, 6.5×10 ⁻⁵)
U9	7.93 s	↓(-0.83, 1.8×10 ⁻³)
Unassigned spectral regions		
1.99-2.01		↑(0.73, 1.6×10 ⁻²)
2.75-2.79		↓(-1.26, 8.8×10 ⁻⁵)
2.89-2.92		↓(-1.06, 2.0×10 ⁻⁴)
3.35-3.37		↑(1.05, 8.6×10 ⁻³)
7.18-7.20		↑(1.12, 5.5×10 ⁻³)
+ 5 unassigned spin systems and 8 spectral regions with <i>p</i> >0.05		

Literature review:

↑ taurine (Foxall *et al.*, 1995, Trump *et al.*, 2006), ↑ TMAO (Foxall *et al.*, 1995, Constantinou *et al.*, 2005), ↑ *myo*-inositol (Foxall *et al.*, 1995), ↑ betaine (Constantinou *et al.*, 2005), ↑NMND (Trump *et al.*, 2006), hippurate, tryptophan, phenylalanine, malate, tyrosine, hydroxybutyrate, *N*-acetylglutamate and proline (direction not specified) (Atzori *et al.*, 2011)

Moreover changes consistent with previous reports are noted with ^c (namely *myo*-inositol and NMND) and changes also found at 2nd T maternal urine of *pre*-PTD women are noted with ^d (3-Me-histidine, 4-OH-hippurate, citrate and creatinine) in Table 6.5. Within this profile, the variations noted for 3-aminoisobutyrate, *cis*-aconitate, ethanolamine, NMND and unassigned regions 2.75-2.79 and 2.89-2.92 ppm, must be acknowledge for their marked statistical significance (*p*-values ranging from 10⁻⁴ to 10⁻⁷).

The urinary metabolic signature of prematurity is shown in Figure 6.8 through the VIP-wheel plot (a, top) and Volcano plot (b, bottom), with metabolites with higher VIP, statistical and biological noted. Out of the metabolites found to vary in premature newborns, the variation of 3-aminoisobutyrate, *cis*-aconitate, ethanolamine, *myo*-inositol, 2-Py, NMND, formate, U11 and U12 clearly stand out in Figure 6.8a,b, due to their corresponding high VIP, effect size and low *p*-value (ranging from 10⁻³ to 10⁻⁷ orders of magnitude). Lastly, one unassigned region (7.58-7.62 ppm) is noted in the VIP-wheel, which was found to be relevant for the PLS-DA separation (given by the high VIP value in the radial scale) but had individual significance below the limit here defined (*p*-value<0.05) supporting the importance of taking into account the whole profile rather than solely integrals with individual significance.

6.3.1.2 Proposed metabolic interpretation of changes related with prematurity

In this cohort, the metabolites found to have higher importance in relation to prematurity were NMND (involved in nucleotide metabolism), 3-aminoisobutyrate (a product of pyrimidines metabolism), formate (involved in purine biosynthesis), *cis*-aconitate (involved in the TCA cycle), ethanolamine and *myo*-inositol (precursors of lung surfactants phosphatidylethanolamine and phosphatidylinositol, respectively) and creatine, creatinine, indoxyl sulphate (IS) and methylguanidine (possibly reflective of altered renal functionnad/or oxidative stress).

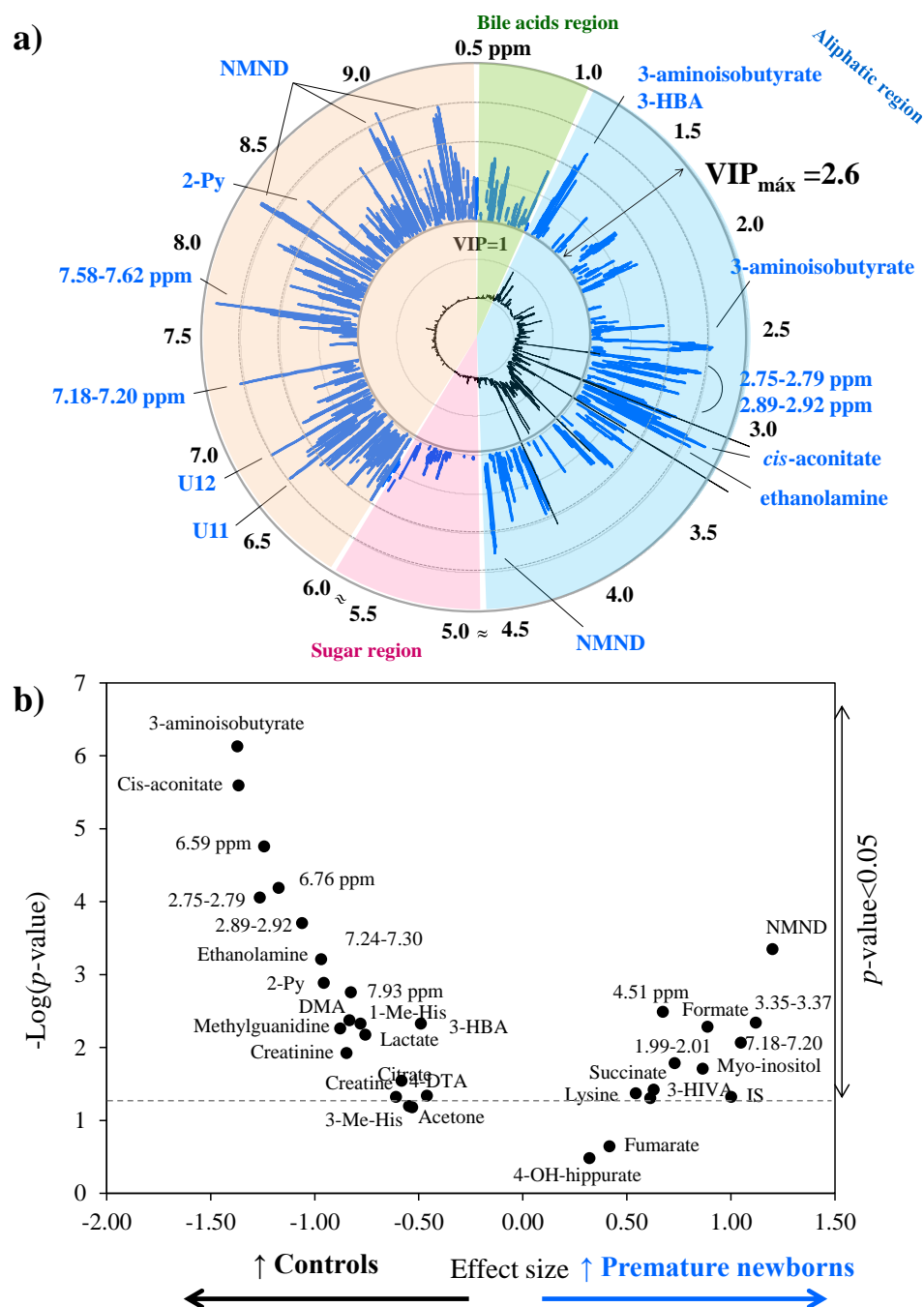


Figure 6.8: a) VIP-wheel representation of the NMR urinary profile of premature newborns. The inner and outer circles represent, respectively, the average urine spectrum of healthy newborn urine (control group) and ppm scale. Main differences are noted. b) Volcano plot of effect size vs. $-\text{Log}(p\text{-value})$ of metabolite/resonances integrals shown in Table 6.4. Dashed line indicates $p\text{-value}=0.05$.

Nucleotide metabolism

NMND is formed from niacin and tryptophan providing pyridines nucleotides to the liver, being one of the end products of nicotinamide metabolism along with 2-Py

(Figure 6.9). Higher NMND had been previously found in premature newborn urine, this variation having been attributed to the high concentration of tryptophan in fetal blood and its regulation after birth (Trump *et al.*, 2006). In this thesis, NMND was found to increase in premature newborns whilst 2-Py decreases. The variation found in NMND can be, as for the previous study (Trump *et al.*, 2006), a reflection of altered tryptophan metabolism after birth. Nevertheless, the concomitant decrease of 2-Py suggests a more profound perturbation in nucleotide metabolism. This hypothesis is further supported by the significant decrease of 3-aminoisobutyrate, the end product of the pyrimidine degradation pathway, particularly of thymine (Figure 6.9), suggesting reduced pyridine degradation. Moreover, formate, an essential metabolite for *de novo* purine biosynthesis, through activation by tetrahydrofolate (THF) (Lamarre *et al.*, 2013), was found to increase in the premature group. The suggestion of altered purines synthesis is consistent with previous studies of umbilical cord blood of premature very low birth weight (VLBW) newborns, which reported altered purines biosynthesis namely comprising an increase of methyladenosine and cytosine and a decrease in amino acid precursors (Alexandre-Gouabau *et al.*, 2013).

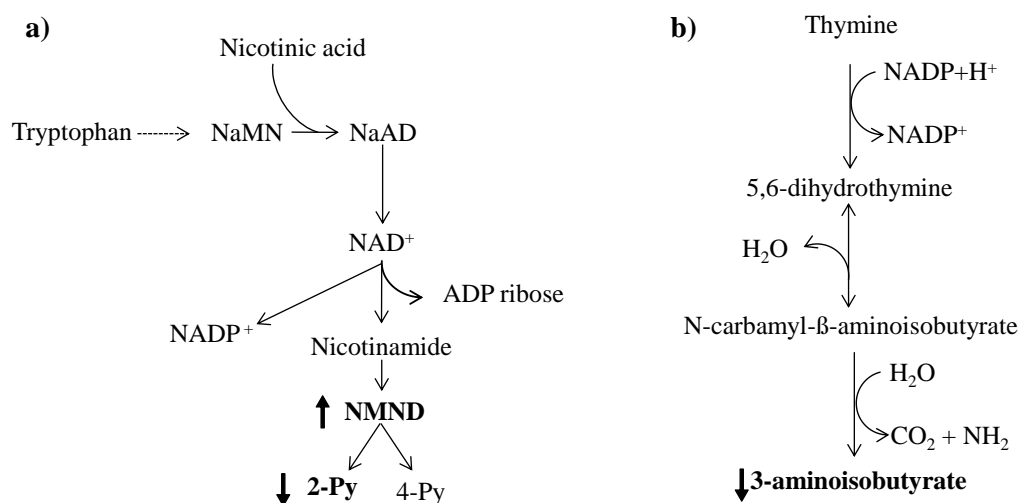


Figure 6.9: a) Metabolic pathway of tryptophan-NAD⁺ biosynthesis and b) thymine degradation pathway. NaMN: nicotinic acid mono nucleotide, NaAD: nicotinic acid adenine dinucleotide, NAD⁺: nicotinamide adenine dinucleotide, NADP: nicotinamide dinucleotide phosphate, ADP: adenosine diphosphate, NMND: *N*-methyl-nicotinamide, 2-Py: *N*-methyl-2-pyridone-5-carboxamide, 4-Py: *N*-methyl-4-pyridone-3-carboxamide. Metabolites detected in the NMR urine spectra are noted in bold. Small arrows indicate variation in the premature group (changes with |effect size|>0.9).

Energy metabolism

Cis-aconitate, an intermediate of the TCA cycle, was found significantly decreased in premature newborns, whilst citrate, fumarate and succinate, also intermediates of the cycle were found increased (Figure 6.10). A derangement in energy metabolism is further supported by a concomitant change in ketone bodies, ketogenic amino acid lysine, 4-DTA (a product of threonine catabolism) and lactate. Previous studies have associated prematurity with enhanced use of fatty acids as a source of energy at the time of birth (Alexandre-Gouabau *et al.*, 2013). In this cohort, ketone bodies 3-HBA and acetone were found decreased in the premature group suggesting their decreased production from fatty acids β -oxidation or higher ketone bodies consumption for energy production. 4-DTA, a product of threonine catabolism was also found to decrease, the excretion of this metabolite having been correlated with 3-HBA in connection with formation of ketone bodies (described in Section 3.3, Chapter 3). Finally, lactate was found to decrease in the premature group, this metabolite being formed from glycolysis under hypoxic conditions. Increased urinary excretion (opposite to the variation found here for the premature group)

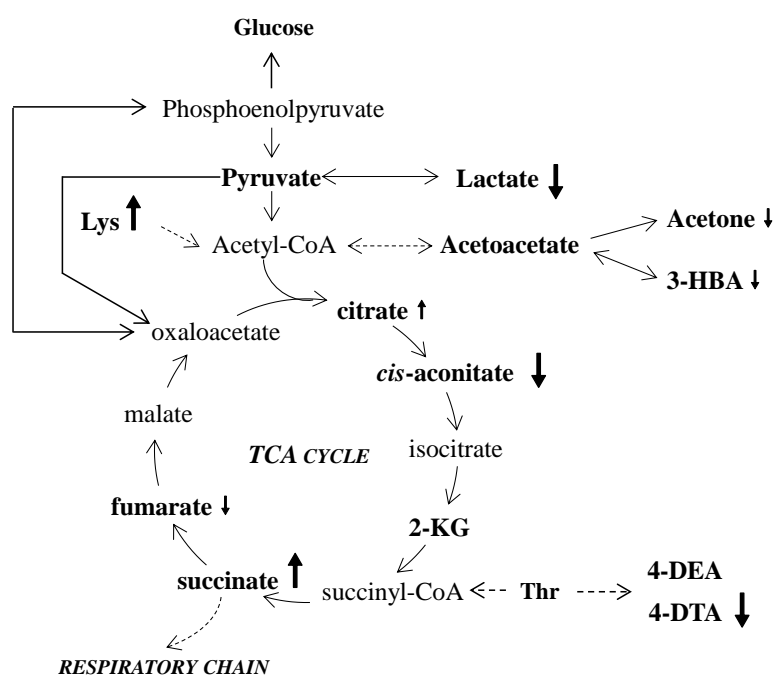


Figure 6.10: Schematic representation of the TCA cycle and ketogenic pathways possibly affected in premature newborns. Amino acids in three-letter code, 3-HBA: 3-hydroxybutyrate, 2-KG: 2-ketoglutarate, 4-DEA: 4-deoxyerythronic acid, 4-DTA: 4-deoxythreonic acid. Metabolites identified in urine NMR spectrum are noted in bold. Small arrows indicate direction of variation in the premature group, with larger arrows indicating metabolite changes with $|\text{effect size}| > 0.5$.

of lactate has been reported in asphyxiated newborns due to hypoxia (Ma *et al.*, 1995), but not in connection to prematurity. The overall results suggest a perturbation in energy metabolism in premature newborns, specifically in the operation of the TCA cycle and ketone bodies production.

Lung surfactant production

Significant variations were also found in this work for ethanolamine and *myo*-inositol, respectively precursors of phosphatidylethanolamine and phosphatidylinositol, both being membrane phospholipids and acting as lung surfactants (Figure 6.11) as reported in (Groenen *et al.*, 2004, Agassandian and Mallampalli, 2013). Apart from these compounds, other lung surfactants include phosphocholine (the most abundant), phosphatidylglycerol (the 2nd most abundant), shingomielyn and phosphatidylserine (less abundants) in the lungs (Agassandian and Mallampalli, 2013). Phosphatidylcholine synthesis increases from week 28th until the end of gestation with peak production at 36th g.w., whilst phosphatidylinositol production increases from the 28th week with maximum values between week 32 and 35 and declining until the end of gestation (Grenache and Gronowski, 2006, Creasy *et al.*, 2009). The last surfactant to be produced is phosphatidylglycerol, appearing after the 36th g.w. until the end of gestation (Grenache and Gronowski, 2006). Regarding phosphatidylethanolamine synthesis no reference gestational age range was found.

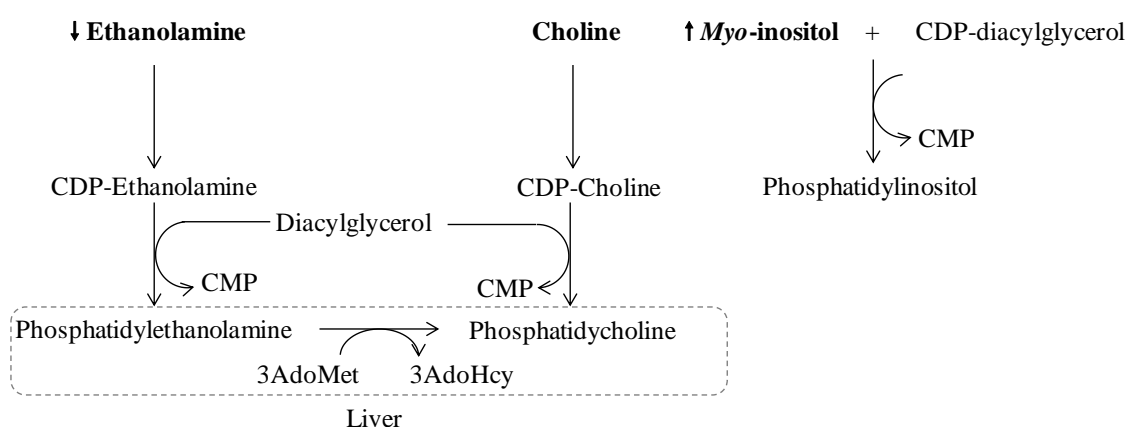


Figure 6.11: Synthesis of phosphatidylethanolamine, phosphatidylcholine and phosphatidylinositol. CDP: cytidine diphosphate, CMP: cytidine monophosphate, AdoMet: *S*-adenosylmethionine, AdoHcy: *S*-adenosylhomocysteine. Metabolites identified in newborn urine are noted in bold. Small arrows indicate direction of variation in the premature group (in this case metabolite changes with |effect size|>0.8).

Increased ethanolamine in urine (ethanolaminuria) of newborns and infants has been associated with altered biosynthesis of membrane phospholipids, brain white matter degeneration, and altered quality/quantity of surfactant synthesis (Cole *et al.*, 1988). The decreased excretion of ethanolamine seen here in the urine of premature newborns suggests its utilization towards phosphatidylethanolamine synthesis, thus indicating an altered profile of surfactants formation at the gestational age range these newborns were born (33-36 g.w.).

Increased *myo*-inositol in urine of premature newborns had been previously reported, the variation having been related to altered renal function as the compound is present in the renal medulla (Foxall *et al.*, 1995). *Myo*-inositol has also been found increased in urine of IUGR newborns possibly due insulin resistance (Dessì *et al.*, 2012). In this cohort, premature babies were born approximately at the gestational age characterized by higher PI formation (32-35 g.w., which further declines until the end of gestation). Thus, increased urinary *myo*-inositol in premature newborns could be a direct reflection of its higher formation (possibly from glucose) to meet the higher demand for PI synthesis and/or a consequence of compromised renal function (the latter hypothesis being discussed below).

Renal function and oxidative stress

Lower creatinine in premature newborn urine has been reported due to immature tubular and vascular structures in the kidney, whilst low creatine was related to its impaired biosynthesis (Lage *et al.*, 2013). High IS was also found in the premature group, this metabolite being known to induce oxidative stress (Dou *et al.*, 2007) and nephrotoxicity (Suhre *et al.*, 2010), possibly reflective of increased oxidative stress in premature newborns, as has been previously reported (Perrone *et al.*, 2007). Increased IS in urine and serum have been related to impaired renal function in adults (Suhre *et al.*, 2010, Huang *et al.*, 2012), once more supporting the possibility of compromised renal function in premature newborns. In this cohort, methylguanidine was found decreased in urine of premature newborns. In addition to IS, methylguanidine is also an uremic toxin formed from creatinine, via creatol, its biosynthesis being enhanced by reactive oxygen species (Marescau *et al.*, 1997). Thus, the variation here noted for methylguanidine may possibly be a direct consequence of decreased creatinine and/or impaired renal function.

Other metabolic effects

Other significant changes were found for 3-HIVA, DMA and 1-methylhistidine, the latter two being related with gut microflora. 3-HIVA has been proposed as a marker of reduced biotin status, having been found increased in maternal serum of women carrying fetuses with trisomy 21 (Bahado-Singh *et al.*, 2013) and in 2nd trimester maternal urine of *pre*-diagnosis GDM (Chapter 5.2). In this cohort, 3-HIVA excretion was found to increase in premature newborns, possibly suggesting altered biotin status although no evidence supporting of this hypothesis was found in literature.

6.3.2 Respiratory depression episode

Breathing complications after birth can lead to the development of perinatal asphyxia, i.e oxygen deprivation, occurring at the time of birth, with impaired gas exchange leading to progressive hypoxemia (low blood oxygen levels) and hypercapnia (high carbon dioxide blood levels). A severe case of asphyxia encompasses the further development of severe acidosis, depressed apgar score at 5th min, neonatal neurological sequelae (i.e. seizures, coma, hypotonia) and multiple organ sequelae (Tax *et al.*, 2013). Other respiratory and asphyxia complications include hypoxic ischemic encephalopathy (HIE), respiratory distress syndrome (RDS) and meconium aspiration syndrome (MAS). HIE occurs due to impaired gas exchange leading to reduced cerebral blood flow, the condition being associated with long-term neurological sequelae, including cerebral palsy, seizures disorders, cognitive delays and motor disabilities (Creasy *et al.*, 2009, Reinke *et al.*, 2013). RDS is defined as respiratory distress within the first hours of life due to a deficiency in lung surfactant, thus often associated with preterm birth (Grenache and Gronowski, 2006), whilst MAS encompasses delivery through meconium stained amniotic fluid, respiratory distress and a characteristic appearance in chest radiographs (Creasy *et al.*, 2009). Neonatal respiratory depression is defined as at least one of the following situations a) newborns requiring resuscitation, b) acute respiratory depression in the first minutes of life after normal cough and breath at birth, c) breathing deterioration after a few minutes with decreasing effort to breathe and d) normal appearance at birth followed by strenuously breathing with signs of asphyxia (Perlman, 1985). The group of samples used in this cohort was defined as newborns affected by a respiratory depression episode

requiring external oxygen supply shortly after birth, with immediate recovery (low Apgar score at 1st minute with recovery showed by normal score at the 10th minute).

6.3.2.1 Unveiling an urinary metabolic signature of respiratory depression

In this work, urine from newborns affected by respiratory depression at birth was collected 1 to 2 days after the episode/birth and compared to controls, in order to unveil the impact of this episode at birth in newborns' metabolism. Other abovementioned respiratory complications were not possible to study as there were no samples available at the time of this thesis. In spite of the low number of samples available for the respiratory depression group ($n=10$), this disorder was found to cause the second highest impact in newborn urine, following prematurity. Visual comparison of the average ^1H NMR spectra of control and respiratory depression groups revealed significant changes in taurine, *myo*-inositol, betaine, glycine and lactose, some being noted in Figure 6.1c, page 212. To confirm these changes and unveil the complete profile of the respiratory depression group, PCA and PLS-DA were applied to the original dataset (not shown) and after variable selection (Figure 6.12a,b).

No group separation was obtained through PCA, although the corresponding score plot unveiled two samples clearly distant from the remaining group and controls, as highlighted in Figure 6.12a. The reviewing of the clinical data from these cases showed that, interestingly, these samples belong to two newborns born from mothers who smoked during pregnancy. It is known that smoking during pregnancy reduces circulating oxygen levels and consecutively its transfer to the fetus, causing chronic fetal hypoxia (Moore *et al.*, 2008). Visual inspection of the urine NMR spectra of these two cases revealed the presence of ethanol in one sample (possibly due to contamination) and high concentrations of unassigned multiplets in the aromatic region (7.06 and 7.43 ppm) in the other sample. The corresponding PLS-DA, revealed a clear separation between groups (Figure 6.12b), with MCCV confirming the robustness and predictive power of the model (CR 91%, sensitivity 96% and specificity 91%, low overlap of Q^2 distributions of true and permuted classes and median Q^2 of 0.500 of the true classes model as shown in Figure 6.12c,d, Table A-III.4, Annex III).

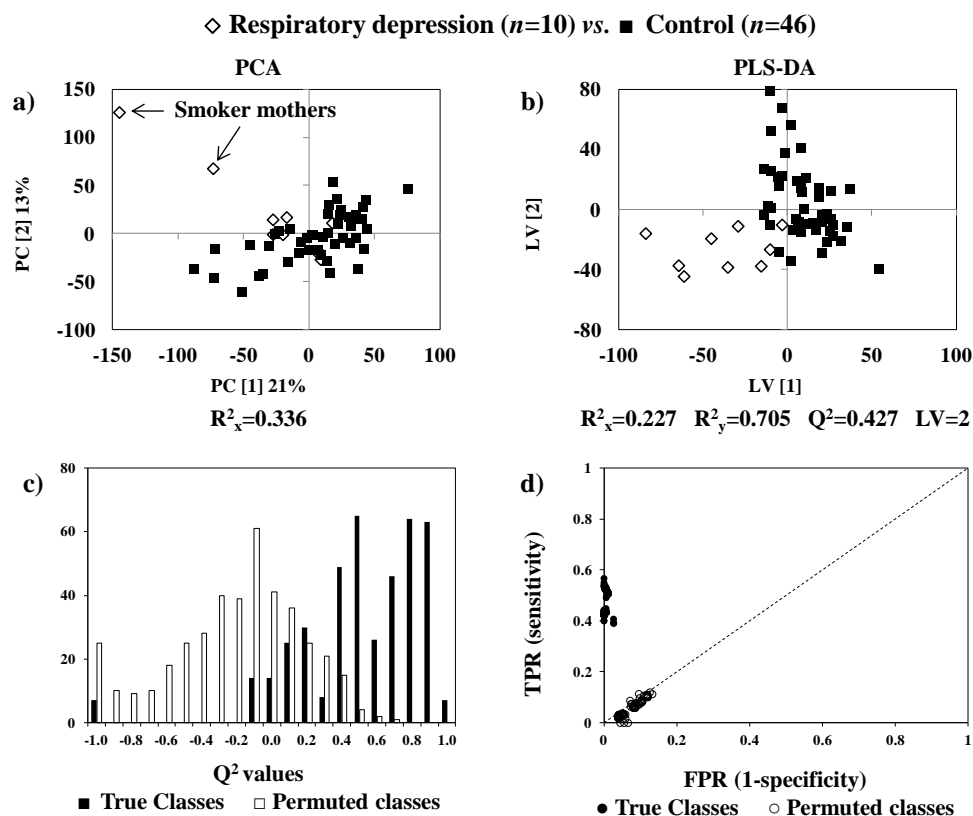


Figure 6.12: a) PCA and b) PLS-DA score plots, c) Q^2 values and d) ROC plot of true and permuted models obtained for urine of newborns with respiratory depression after birth ($n=10$) compared to controls ($n=46$). TPR: true positive rate, FPR: false negative rate.

The analysis of the subset of variables on the basis of the separation seen through PLS-DA enabled the identification of a metabolic signature characteristic of the respiratory depression group (Table 6.6), along with a brief review of relevant literature for comparison. The unveiled signature comprises variations in 21 metabolites (4-DEA, betaine, creatine, glycine, *myo*-inositol and taurine having p -value <0.05), 6 unassigned spin systems (singlets at 2.05, 2.06, 2.26, 2.81 and 8.39 ppm) and 5 unassigned spectral regions with p -value <0.05 .

Metabolites/resonances of this signature with high VIP are further shown by a VIP-wheel representation (Figure 6.13a), whilst integrals with high biological and statistical relevance are highlighted through the Volcano plot shown in Figure 6.13b (dashed line indicates limit of statistical significance, p -value <0.05). The inspection of the VIP-wheel and Volcano plot revealed that metabolites with higher contribution for the respiratory depression signature were glycine, betaine (the latter two metabolites having the lowest

Table 6.6: Metabolite/resonance changes in newborn urine of respiratory depression group. ^a Chemical shifts selected by variables selection; TOCSY or STOCYSY correlated peaks in square brackets; s: singlet, d: doublet, t: triplet, m: multiplet. Peaks possibly related to diet and/or gut microflora. Ui: unassigned compound i by order of appearance and following Table 6.2, Table 6.4 and Table 6.5. Abbreviations defined in Table 6.2, Table 6.4 and Table 6.5. Only *p*-values < 0.05 are shown.

Respiratory Depression (n=10) vs. Control (n=46)		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
1-Me-histidine ^b	7.78 s	↓(-0.27)
2-KG	2.45 t, 3.01 t	↓(-0.17)
3-HBA	2.31 m, 2.41 m, 4.15 m	↓(-0.65)
3-Me-histidine ^b	7.15 s	↓(-0.08)
4-DEA	1.11 d	↓(-0.67, 4.43×10 ⁻²)
4-DTA	1.23 d	↓(-0.49)
Betaine	3.26 s, 3.90 s	↑(1.10, 6.33×10 ⁻³)
Choline ^b	3.20 s	↑(0.65)
Creatine	3.04 s, 3.93 s	↑(0.95, 1.54×10 ⁻²)
DMA	2.72 s	↓(-0.65)
Formate	8.46 s	↑(0.72)
Glycine	3.57 s	↑(1.21, 1.65×10 ⁻³)
Hypoxanthine	8.20 s, 8.22 s	↓(-0.58)
IS	7.70 d	↑(0.24)
Lactose ^b	3.66 m, 3.79 m, 3.94 m, 4.46 d	↑(0.93)
Lysine	1.73 m, 1.92 m	↓(-0.73)
Myo-inositol	3.26 t, 3.55 m, 3.64 t, 4.07 t	↑(1.54, 1.18×10 ⁻²)
Taurine	3.43 t, 3.26 t	↑(0.96, 1.10×10 ⁻²)
Threonine	4.26 dd	↓(-0.60)
Trigonelline ^b	4.44 s	↓(-0.04)
Xylose ^b	5.21 d	↓(-0.15)
U13	2.05 s	↑(1.00, 1.54×10 ⁻²)
U2	2.06 s	↓(-0.72, 2.1×10 ⁻²)
U14	2.26 s	↓(-0.82, 3.97×10 ⁻²)
U15	2.81 s	↑(0.88, 2.23 ×10 ⁻²)
U16	2.90 s [3.94 s]	↑(0.90, 1.27×10 ⁻²)
U17	8.39 s [6.08 d]	↑(1.10, 4.92×10 ⁻²)
Unassigned spectral regions		
2.08-2.09		↑(1.00, 1.11×10 ⁻²)
2.10-2.12		↓(-0.85, 3.76×10 ⁻²)
2.48-2.52		↓(-0.79, 4.19×10 ⁻²)
6.61-6.62		↓(-1.03, 3.36×10 ⁻²)
6.63-6.65		↓(-1.00, 3.36×10 ⁻²)
+8 unassigned spin systems and 16 spectral regions with <i>p</i> >0.05		

Literature review of studies in newborn urine:

Perinatal asphyxia: ↑lactate, ↓citrate, succinate, 2-KG, DMA, DMG, glucose (Ma *et al.*, 1995); ↑ ethylmalonate, 3-hydroxy-methylglutarate, 2-hydroxyglutarate, 2-KG, glutarate, methylmalonate, 3-hydroxybutyrate, orotate (Chu *et al.*, 2006)

p-values of this profile, 10⁻³ orders of magnitude), myo-inositol (this metabolite having the highest effect size of this profile, 1.54), taurine and 4-DEA. Additionally, unassigned spin systems U2, U13, U14, U15 and U17 were also found to have a significant contribution for the respiratory depression signature as given by the high VIP value, effect size and

statistical significance. Lastly, several unassigned regions from 6.09-6.65, 7.18-7.20 and 8.13-8.15 were found to be relevant for the urinary profile here unveiled, supported by their high VIP value in the VIP-wheel radial scale although, individually, these regions were not found to be statistically significant

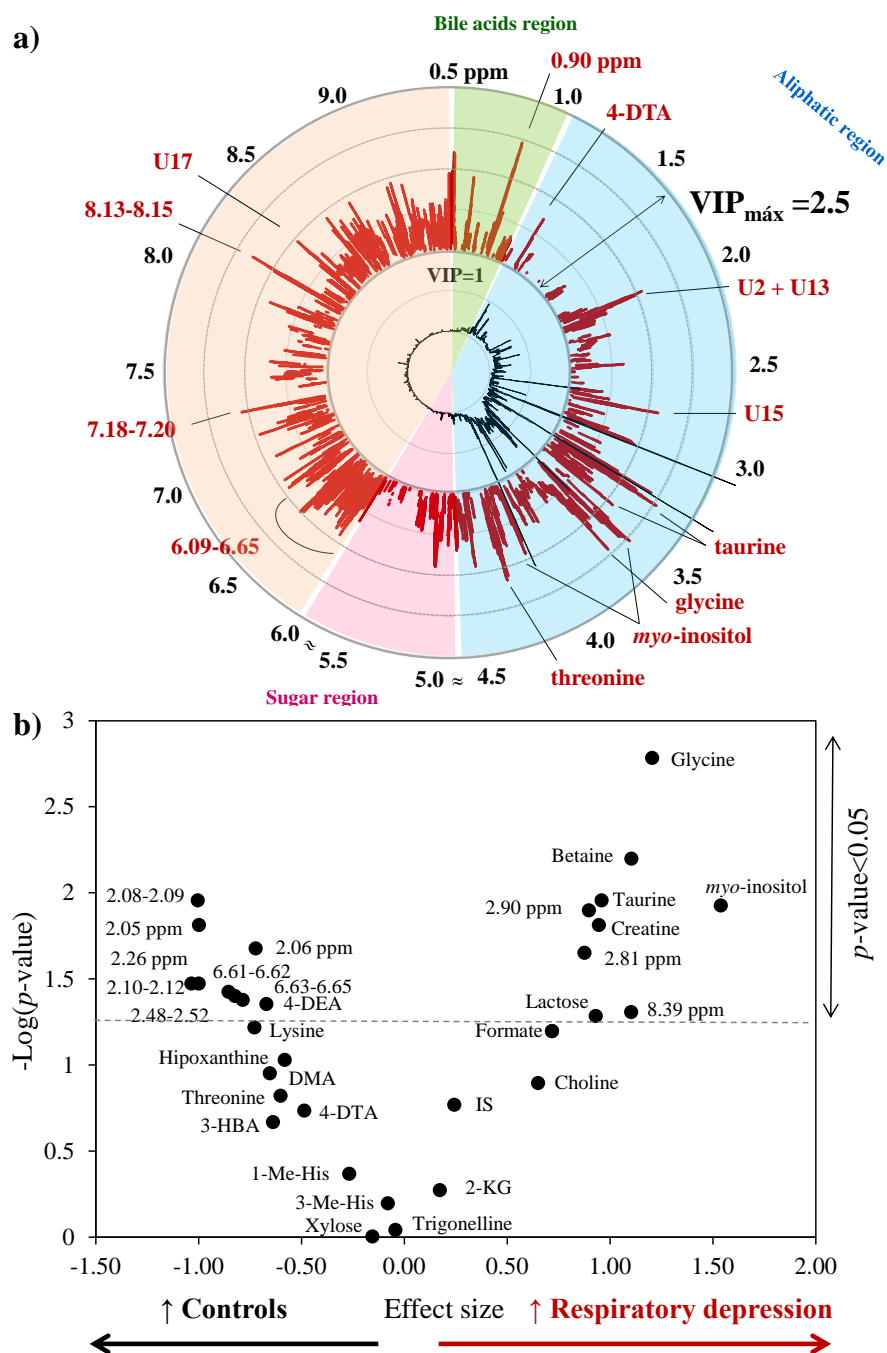


Figure 6.13: a) VIP-wheel representation of the NMR urinary profile of the respiratory depression group. The inner and outer circles represent, respectively, the average urine spectrum of healthy newborn urine (control group) and corresponding ppm scale. Main differences are noted. b) Volcano plot of effect size vs. $-\text{Log}(p\text{-value})$ of metabolite/resonances integrals shown in Table 6.6.

6.3.2.2 Proposed metabolic interpretation of changes related with respiratory depression

This cohort includes urine from newborns affected by a respiratory depression episode at birth, with depressed Apgar score at 1st min, but without other clinical signs of asphyxia (namely hypoxemia, hypercapnia, acidosis, neonatal neurological and other organ sequelae). Still, a parallelism with neonatal asphyxia may be drawn carefully when interpreting the results, bearing in mind that neonatal asphyxia implies a more severe case of oxygen deprivation and comprises other sequelae.

In this study, the variation of glycine, betaine, taurine, creatine and 4-DEA were found to have the highest individual biological and statistical significance. Increased glycine, an excitatory neurotransmitter, has been reported in cerebral spinal fluid (CSF) (Gucuyener *et al.*, 1999, Roldan *et al.*, 1999) and brain (Malik *et al.*, 2002) of neonates with HIE. It has been suggested that perinatal asphyxia modifies cerebral synthesis of neurotransmitters, with enhanced production of glycine (Roldan *et al.*, 1999). The results here presented suggest that this enhanced requirement for glycine could explain the higher production and urinary excretion of this amino acid found in the respiratory depression group. Glycine is formed from DMG in the methionine metabolism. Glycine is also converted to guanidoacetate (GAA), and further methylated to creatine with formation of homocysteine, the latter being transsulfurated forming taurine. A schematic representation of the metabolic relations between these metabolites is shown in Figure 6.14, with indication of metabolites detected in newborn urine in bold and metabolites found to vary in this cohort with short arrows. Increased glycine, betaine, choline, creatine and taurine here found suggest a profound perturbation in methionine metabolism following the respiratory depression episode, probably for the replenishing of the increased glycine requirement towards enhanced neurotransmitter formation.

Another relevant change in this group is the increase in *myo*-inositol levels. Higher *myo*-inositol levels have been reported in the brain of HIE newborns through *in vivo* NMR measurements (Fan *et al.*, 2003), this finding being related to gliosis (proliferation of astrocytes after central nervous system, CNS, injury) and poor myelination, as *myo*-inositol has been labeled as a breakdown product of myelin (Fan *et al.*, 2003). This compound is also a precursor of lung surfactant phosphatidylinositol (PI). Hence, higher *myo*-inositol levels can also be related to changes in lung surfactant, possibly leading and explaining the

development of the respiratory depression episode. However, the origin of the increase in *myo*-inositol found in this group differs from that noted for premature newborns (section 6.3.1) as, in the latter group, this variation could be explained by the gestational age at birth, which coincided with the peak of PI formation (32-35 g.w. declining until the end of gestation). This is not the case of the respiratory depression group (GA range at birth for this group is 36-40 g.w.). Thus, it is possible that the increase in *myo*-inositol excretion originates from a perturbation in surfactant production (unexpected based on GA at birth) hence causing the episode, and/or altered myelination, the latter being associated with a possible brain insult caused by the oxygen deprivation.

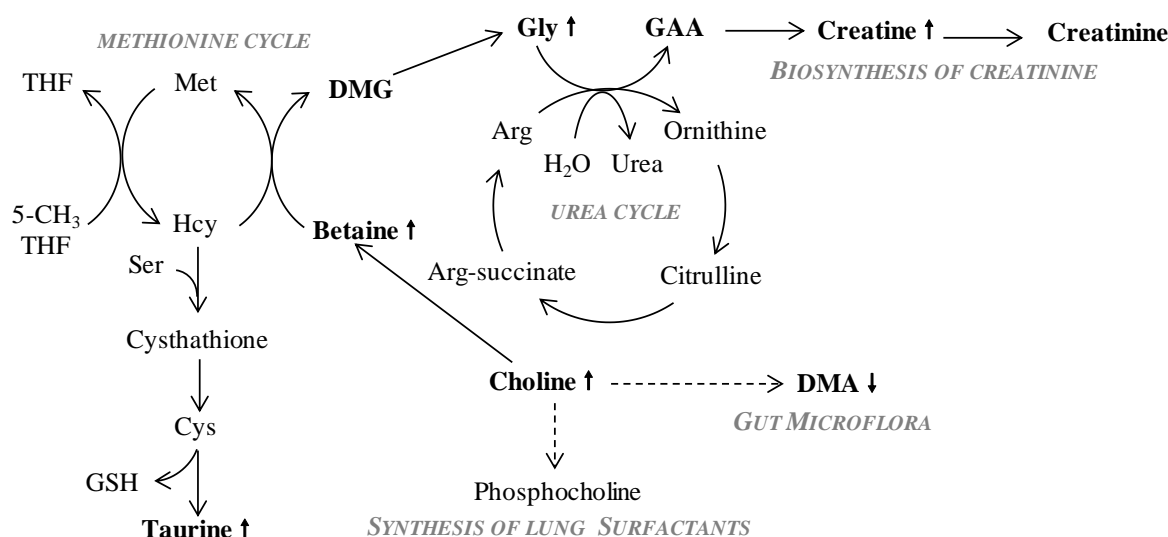


Figure 6.14: Representation of the metabolic relationship between methionine metabolism, creatine biosynthesis, urea cycle and synthesis of lung surfactants (adapted from (Brosnan and Brosnan, 2010, Dasarathy *et al.*, 2010)). Metabolites identified in newborn urine are noted in bold. Small arrows direction of variation in the respiratory depression group.

6.3.3 Large for Gestational Age (LGA)

Newborns are considered large for gestational age, LGA, when their birth weight is above the 90% percentile for the corresponding gestational age at birth. Being born LGA poses higher risk for birth injuries, including cervical and brachial plexus injuries, phrenic nerve damage, with paralysis of the diaphragm, fractured clavicles, cephalohematomas, subdural hematomas and ecchymoses of the head and face (Kliegman *et al.*, 2011). Additionally, LGA newborns have higher risk of developing hypoglycemia and

polycythemia (higher concentration of red blood cells) and also higher incidence of congenital anomalies, particularly of the heart (Kliegman *et al.*, 2011). Long-term effects such as intellectual and developmental retardation are also more common in these newborns (Kliegman *et al.*, 2011). LGA is often caused by abnormalities in maternal-fetal placental circulation and maternal glucose metabolism, this growth perturbation being common in newborns of obese women and GDM-affected pregnancies due to fetal hyperglycemia and hyperinsulinemia (Hod *et al.*, 2008, Kliegman *et al.*, 2011).

6.3.3.1 Unveiling an urinary metabolic signature of LGA

In this study, urine from LGA newborns was compared with that of controls i.e. healthy newborns with birth weight adequate for gestational age (AGA), more information regarding samples metadata being shown in Table 2.5 of Experimental Section. For the LGA group only 2 out of the 18 samples were obtained from GDM-affected pregnancies (LGA being a common complication of GDM), thus the unveiled urinary profile is likely reflective exclusively of LGA and not of GDM. PCA and PLS-DA were applied to the original dataset and after variable selection and validated through MCCV. Although no group separation was obtained through PCA (Figure 6.15a), the PLS-DA suggested a separation trend between LGA and controls, as shown in Figure 6.15b. MCCV suggested a relatively robust model, seen through ROC plot analysis and confusion matrices, which yielded CR 84%, sensitivity 76%, specificity 86%, and predictive power supported by Q^2 distribution of true and permuted classes and median Q^2 of 0.455 for true classes (Figure 6.15c,d, and Table A-II.4, Annex 2). Regarding the two LGA located on the positive direction of LV[2] and noted in Figure 6.15b, visual inspection of the corresponding NMR spectra revealed higher levels of the unassigned region 0.70-1.07 ppm in one, and higher *myo*-inositol in the other sample, all metabolites found relevant for the LGA urinary signature.

The subset of variables on the basis of the PLS-DA separation unraveled a urinary metabolic signature of LGA (Table 6.7). At the end of this table, a summary of relevant changes in newborn urine in cases of other fetal growth perturbations (namely intrauterine growth restriction-IUGR) reported in literature is shown. The LGA profile here unveiled comprised changes in 13 metabolites, including 4 amino acids and derivatives, 4 organic

acids, 2 products of nucleotide metabolism, 1 metabolite deriving from the gut microflora and 2 others (*myo*-inositol and creatine) along with 13 unassigned regions with $p < 0.05$, mostly at 0.5-1 ppm and 6-7 ppm. Since the LGA group was found to be unbalanced in terms of gender, comprising 14 males and 4 females newborns, a possible confounding contribution of gender needed to be taken in to account (changes previously listed in Table 6.4). This means that the increased excretion noted for creatine and hippurate along with decreased excretion of 2-KG and fumarate (noted with * in Table 6.7) may arise from the unbalance of gender rather than the LGA condition.

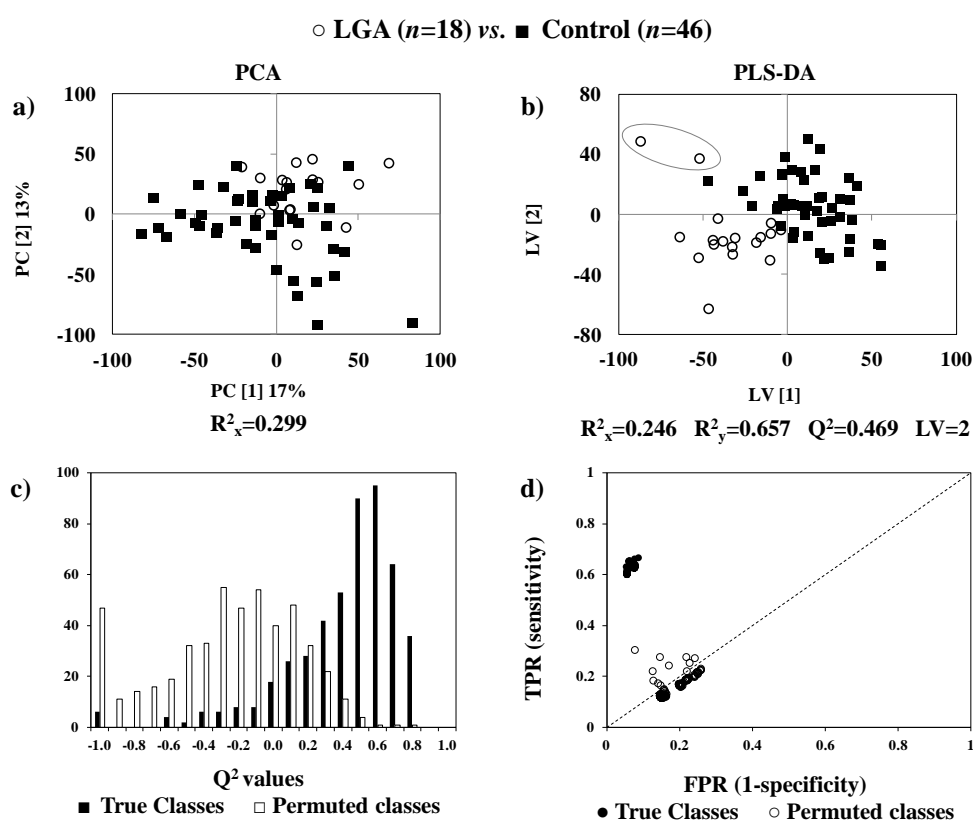


Figure 6.15: a) PCA and b) PLS-DA score plot, c) Q^2 values and d) ROC plot of true and permuted models obtained for urine of newborns born large for gestational age (LGA, $n=18$) compared to controls ($n=46$). TPR: true positive rate, FPR: false negative rate.

Furthermore, metabolites/resonances relevant for the profile were highlighted through the VIP-wheel representation and Volcano plot shown in Figure 6.16a,b. Through the inspection of Figure 6.16 it becomes clear that 2-Py, DMG and *myo*-inositol, along with several unassigned regions mainly at 0.5-1.0 (bile acids region (Trump *et al.*, 2006,

Duarte *et al.*, 2009)) and 6-7 (purines and pyrimidines region (Wevers *et al.*, 1999)) ppm have the highest individual biological and statistical significance (p -values ranging from 10^{-2} to 10^{-3}) the same metabolites/regions having high VIP. One exception however is that of hippurate and unassigned region 1.84-1.90 ppm which were found to be relevant for the separation (given their high VIP value) but not individually significant (p -value >0.05).

Table 6.7: Metabolite/resonance changes in newborn urine of LGA group. ^a Chemical shifts selected by variables selection; s: singlet, d: doublet, dd: doublet of doublets, t: triplet, m: multiplet. ^b Peaks possibly related to diet and/or gut microflora. *Possible contribution of gender. Ui: unassigned compound i by order of appearance and following Table 6.2, Table 6.4, Table 6.5 and Table 6.6. DMG: dimethylglycine, other abbreviations defined in following Table 6.2, Table 6.4, Table 6.5 and Table 6.6. Only p -values <0.05 are shown.

LGA (n=18) vs. Control (n=46)		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, p -value)
2-KG*	2.45 t, 3.01 t	↓(-0.18)
2-Py	6.66 d, 7.97 dd, 8.33 d	↑(1.01, 9.6×10^{-3})
Alanine	1.49 d	↑(0.10)
Citrate	2.54 d, 2.69 d	↓(-0.40)
Creatine*	3.04 s, 3.93 s	↑(0.34)
DMG	2.93 s	↑(0.77, 6.3×10^{-3})
Fumarate*	6.53 s	↓(-0.51)
Glycine	3.57 s	↓(-0.43)
Hippurate ^{b,*}	3.97 d, 7.56 t, 7.64 t, 7.83 d	↑(0.56)
Myo-inositol	3.29 t, 3.55 t, 3.63 t, 4.08 t	↓(-0.67, 1.7×10^{-2})
NMND	8.18 m, 8.90 d, 8.97 d, 9.28 s	↑(0.58)
Succinate	2.41 s	↑(0.59)
Taurine	3.43 t	↑(0.14)
Unassigned spectral regions		
0.5-0.55		↑(0.76, 1.4×10^{-3})
0.57-0.60		↑(0.56, 3.3×10^{-3})
0.64-0.70		↑(0.70, 1.6×10^{-3})
0.77-0.87		↑(0.79, 2.8×10^{-2})
0.89-0.94		↑(1.04, 2.1×10^{-3})
3.35-3.38		↓(-0.66, 1.4×10^{-2})
6.17-6.19		↓(-0.49, 1.4×10^{-2})
6.35-6.36		↓(-0.48, 3.0×10^{-2})
6.39-6.41		↓(-0.65, 8.0×10^{-3})
6.43-6.45		↓(-0.87, 1.3×10^{-3})
6.46-6.47		↓(-0.80, 4.4×10^{-3})
6.68-6.71		↑(0.55, 3.8×10^{-2})
8.01-8.07		↑(1.30, 2.4×10^{-2})
+ 12 unassigned spin systems and 15 spectral regions with $p>0.05$		
Literature review: IUGR newborn urine: ↑ myo-inositol, creatine, creatinine and sarcosine (Dessi <i>et al.</i> , 2011)		

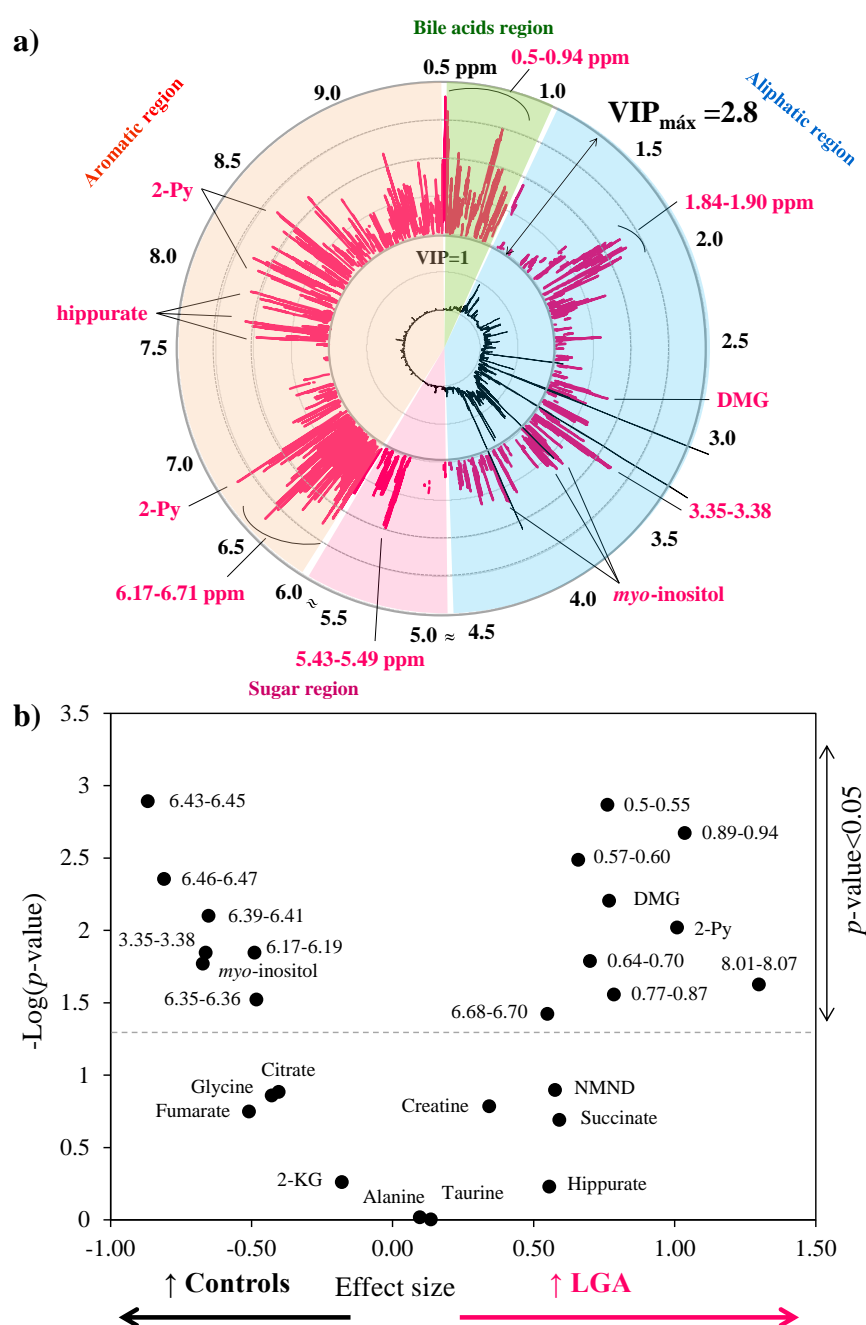


Figure 6.16: a) VIP-wheel representation of the NMR urinary profile of newborns born LGA (inner and outer circles represent, respectively, the average urine spectrum of healthy newborn urine (control group) and corresponding ppm scale). b) Volcano plot of effect size vs. $-\log(p\text{-value})$ of metabolite/resonances integrals summarized in Table 6.7. Dashed line indicates $p\text{-value} = 0.05$.

6.3.3.2 Proposed metabolic interpretation of changes related with LGA

To our knowledge, this is the first metabolomic report addressing the effect of being born LGA, since previous studies have only focused on IUGR (Dessì *et al.*, 2011,

Favretto *et al.*, 2012), SGA (Horgan *et al.*, 2011), LBW (Ivorra *et al.*, 2012) and VLBW (Alexandre-Gouabau *et al.*, 2013).

In this cohort, urine metabolites with higher statistical and biological significance include *myo*-inositol, DMG and 2-Py. *Myo*-inositol was found to decrease in the LGA group in this study, the plasma concentration of this metabolite having been found to correlate negatively with birth weight in an animal model of LBW vs. normal weight (Nissen *et al.*, 2011). Moreover, a previous work reported increased *myo*-inositol levels in urine of IUGR newborns (inverse growth perturbation to LGA) due to reduced insulin secretion (Dessi *et al.*, 2011). *Myo*-inositol is synthesized from glucose and it has been reported to have insulin-mimetic properties stimulating cellular glucose uptake and glycogen synthesis (Croze and Soulage, 2013). It is also a precursor of lung surfactant phosphatidylinositol (PI) (Groenen *et al.*, 2004, Agassandian and Mallampalli, 2013) and a breakdown product of myelin after brain insult (Fan *et al.*, 2003). However, alterations in lung surfactant production and/or brain insults are unlikely causes of the *myo*-inositol variation encountered in this cohort as all newborns were born at approximately the same gestational age (median GA 39 g.w. for controls and LGA), after the peak of PI formation (week 32 to 35 (Grenache and Gronowski, 2006, Creasy *et al.*, 2009)) and no clinical evidence of brain injury was reported in either LGA or control groups. Thus, the significant decrease of urinary *myo*-inositol in LGA newborns can reflect altered glucose handling leading to enhanced fetal growth (glucose being the main fuel for fetal growth).

Other relevant findings for the LGA group comprise higher urinary DMG and 2-Py levels, these being involved in the methionine cycle and nucleotides metabolism, respectively. DMG has been found increased in umbilical cord blood plasma and correlated with higher birth weight, in a LBW vs. normal weight cohort, due to alterations in choline metabolism (Hogeveen *et al.*, 2013). Moreover, it has been suggested that alterations in methionine metabolism could impact on purine and pyrimidine metabolism, DNA synthesis and/or cell proliferation, and affect fetal growth (Kalhan and Marczewski, 2012). Indeed, the results found in this work are consistent with a perturbation in methionine metabolism towards enhanced production of DMG. Moreover, the tandem increase of 2-Py and NMND was found in LGA newborns, the latter metabolites being end products of nucleotides metabolism, specifically the tryptophan-NAD⁺ pathway that

supplies pyridines nucleotides to the liver (see Figure 6.9). Hence, higher excretion of 2-Py and NMND suggest a further perturbation in nucleotides metabolism.

Regarding the significant contribution of several unassigned spectral regions that characterize the LGA urinary profile, emphasis must be given to the 0.5-1.0 ppm region, which typically contains resonances from bile acids (Trump *et al.*, 2006, Duarte *et al.*, 2009), and is significantly increased in LGA newborns. However, no possible explanation was found in literature linking altered urinary bile acids excretion and LGA. Also, the 6-6.5 ppm range was found significantly decreased in the LGA group, this region typically containing resonances of purines and pyrimidines (Wevers *et al.*, 1999), consistently with above suggested perturbation in nucleotides metabolism.

6.3.4 Malformations

6.3.4.1 Unveiling an urinary metabolic signature of malformations

This study comprised urine from 9 newborns born with a malformation, only 1 sample being a major malformation (corresponding to a case of *situs inversus* with dextrocardia) and 8 minor malformations (comprising 5 cardiac, 2 soft tissues and 1 urogenital). This group was compared to controls by applying PCA and PLS-DA to the original dataset and after variable selection (the best model being obtained with the reduced matrix) and confirmed through MCCV (Table A-III.4, Annex III). No separation was obtained by PCA, as shown by the score plot in Figure 6.17a, whilst PLS-DA revealed a separation trend between malformations and controls, as shown in Figure 6.17b. The PLS-DA model was found to be relatively robust with CR of 89%, sensitivity of 90% and specificity of 73% and relatively good predictive power, given by Q^2 distribution (median Q^2 of 0.391), as shown in Figure 6.17c,d. Regarding the outlier sample noted in both PCA and PLS-DA score plots, this sample corresponding to a newborn diagnosed with of a cardiac malformation, specifically a interventricular communication (IVC) case, the mother being affected by chronic bronchitis. Additionally, visual inspection of the corresponding urine spectra showed high levels of galactose.

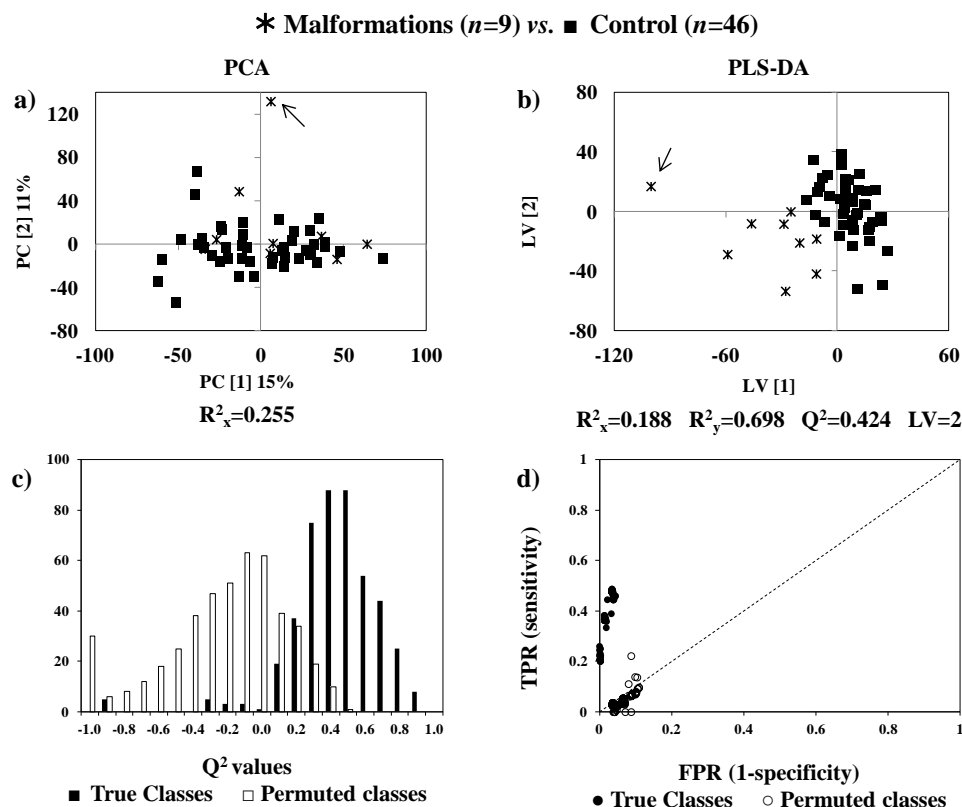


Figure 6.17: a) PCA and b) PLS-DA score plot, c) Q^2 values and d) ROC plot of true and permuted models obtained for newborn urine corresponding to the malformations group ($n=9$) compared to controls ($n=46$). TPR: true positive rate, FPR: false negative rate. Arrows note one outlier sample within the malformation group.

The urinary profile of the malformations group was retrieved considering the selected subset of variables used (Table 6.8), and the overall signature was found to comprise changes in 22 metabolites, including 5 amino acids and derivatives, 4 organic acids, 3 products of nucleotide metabolism, 2 ketone bodies and 8 metabolites deriving from diet and/or gut microflora. Out of the registered profile, 2-Py, citrate, IS and succinate were found statistically significant, 1 unassigned singlet at 4.53 ppm (U18) and 1 unassigned spectral region (8.40-8.42 ppm) with $p < 0.05$. An interesting finding of this profile was that increased NMND and decreased succinate were found in newborn urine commonly with previous findings in 2nd trimester maternal urine of women carrying babies with malformations (Section 4.1.1), these being noted with ^c in Table 6.8. Moreover, 7 out of the 9 malformation cases were born through CS, so a possible contribution of this confounder must to be acknowledged (changes related to delivery mode shown in Table 6.2). This means that the variation noted for 4-HPA, IS, lactose and PAG (all metabolites

possibly deriving from the gut microflora and noted with * in Table 6.8), may have a concomitant contribution from the effect of delivery mode rather than the malformation itself.

Table 6.8: Metabolite/resonance changes in newborn urine of the malformation group. ^a Chemical shifts selected by variables selection; s: singlet, d: doublet, dd: doublet of doublets, t: triplet, m: multiplet. ^b Peaks possibly related to diet and/or gut microflora. ^c Metabolites found to vary commonly with maternal urine of women carrying babies with malformations (Section 4.1.1.). *Possible contribution of delivery mode. Ui: unassigned compound i by order of appearance and following Table 6.2, Table 6.4, Table 6.5, Table 6.6 and Table 6.7. Abbreviations defined in following Table 6.2, Table 6.4, Table 6.5, Table 6.6 and Table 6.7. Only *p*-values<0.05 are shown.

Malformations (n=9) vs. Control (n=46)		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, <i>p</i> -value)
2-Py	6.66 d, 7.97 dd 8.33 d	↓(-0.82, 5.4×10 ⁻²)
3-aminoisobutyrate	1.19 d, 2.61 m, 3.06 dd	↓(-0.64)
3-HBA	1.20 m, 2.31m, 2.41 m, 4.16 m	↓(-0.55)
4-DEA	1.11 d	↑(0.08)
4-HPA ^{b, *}	6.86 d	↓(-0.05)
Acetone	2.23 s	↓(-0.50)
Cis-aconitate	3.12 d	↓(-0.46)
Citrate	2.54 d, 2.67 d	↓(-0.56, 4.3×10 ⁻²)
Choline ^b	3.20 s	↓(-0.19)
DMG	2.93 s, 3.72 s	↑(0.74)
Fumarate	6.53 s	↓(-0.68)
Glutamine	2.11 m, 2.47 m	↑(0.70)
Hippurate ^b	3.97 d, 7.56 t, 7.64 t	↑(0.17)
IS ^{b, *}	7.28 m, 7.50 d, 7.70 d	↓(-0.78, 3.6×10 ⁻²)
Lactose ^{b, *}	3.28 t, 4.46 d	↑(0.39)
Lysine	1.73 m	↓(-0.13)
NMND ^c	8.90 d, 8.97 d, 9.28 s	↑(0.51)
PAG ^{b, *}	7.36 t, 7.43 t	↓(-0.56)
Succinate ^c	2.41 s	↓(-0.93, 3.7×10 ⁻³)
Tyrosine	6.90 d	↑(0.41)
Trigonelline ^b	8.84 br, 9.12 s	↑(0.16)
Xylose ^b	4.59 d, 5.22 d	↑(0.59)
U18	4.53 s [1.84 m]	↑(0.77, 4.5×10 ⁻²)
Unassigned spectral regions		
8.40-8.42		↑(0.99, 2.8×10 ⁻²)
+8 unassigned spin systems and 14 spectral regions with <i>p</i> >0.05		

The above mentioned urinary profile of carrying malformations was also plotted by means of a VIP-wheel and Volcano plot (respectively in Figure 6.18a,b). An immediate observation of the VIP-wheel is the high importance of the aromatic region (shaded in orange) for the PLS-DA group separation given by the high VIP values of almost the whole region. Out of the metabolites/peaks resonating at this region, only 2-Py, IS and the

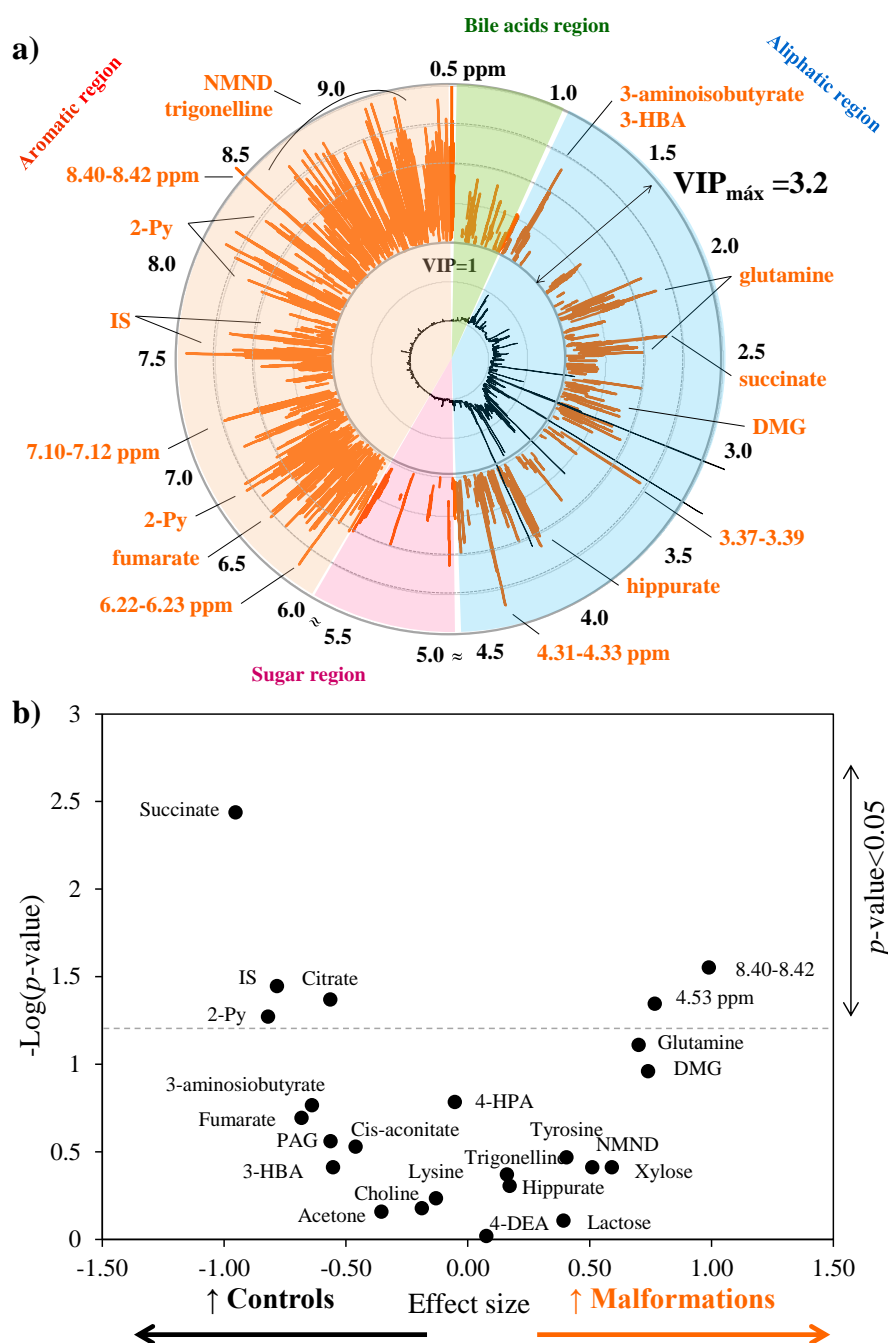


Figure 6.18: a) VIP-wheel representation of the NMR urinary profile of newborns born LGA (inner and outer circles represent, respectively, the average spectrum of the control group and ppm scale) Volcano plot of effect size vs. $-\log(p\text{-value})$ of metabolite/resonances integrals listed in Table 6.8. Dashed line indicates $p\text{-value}=0.05$.

unassigned region 8.40-8.42 were found to have individual statistical significance ($p\text{-value}<0.05$), while others also play an important role in the definition of this profile (namely 4-HPA, hippurate, PAG, NMND, tyrosine, trigonelline, and unassigned regions

6.22-6.23 and 7.10-7.12 ppm, p -value>0.05) and must therefore be acknowledged. However, as previously mentioned, 7 out of 9 malformation cases were born through CS, this delivery mode being related with changes in the urinary excretion of metabolites derived from the gut microflora (these typically resonating in the aromatic region). Hence, the increased importance of the aromatic region in this profile (particularly that of 4-HPA, IS and PAG resonances) may be hampered by the concomitant effect of delivery mode rather than solely reflecting the effect of carrying a malformation. Regarding other metabolites with high contribution for the groups' separation, the VIP-wheel representation and volcano plot showed that succinate, citrate and U18 had high biological and statistical relevance for the profile. Moreover, 3-aminoisobutyrate, glutamine, DMG, fumarate, acetone, 3.-HBA (all with effect size>|0.5|) and unassigned regions 3.37-3.39 and 4.31-4.33 were also found to be relevant for the PLS-DA separation (noted in the VIP-wheel due to their high VIP values) although without individual statistical significance.

6.3.4.2 Proposed metabolic interpretation of changes related with malformations

In spite of the exploratory nature of this model, given the low number of samples available at this stage, a tentative metabolic picture accompanying the occurrence of malformations (mostly minor cases) can be drawn. Metabolites with higher biological and statistical relevance include succinate, IS, citrate and 2-Py, along with the unassigned singlet at 4.53 ppm and the spectral region of 8.40-8.42 ppm. Succinate was found to be significantly decreased in the malformation group, possibly reflecting its underuse in the respiratory chain. This finding is consistent with previous reports that found succinate to increase in AF (Graça *et al.*, 2010) and decrease in maternal urine of women carrying malformed fetuses (Chapter 4, Section 4.1.1). Succinate is also an intermediate of the TCA cycle along with citrate, *cis*-aconitate and fumarate, all being decreased in this cohort, consistently with a slowing down of this pathway. Decreased IS, previously mentioned to induce oxidative stress (Dou *et al.*, 2007) and nephrotoxicity (Suhre *et al.*, 2010), was also found in the malformation group. It would be expected that increased oxidative stress in the pathological group was reflected by an increase in IS, as suggested for maternal urine of CNS malformations cases, described in Section 4.1.1, thus posing an apparent

contradiction. Still, IS was found to decrease in urine of newborns delivered through CS, as discussed in section 6.2.1, possibly due to labor-related stress. Given that 7 out of the 9 malformation cases were born through CS, it is possible that the decreased IS levels found in the malformation group is a reflection of delivery mode effects rather than the malformation itself. Hence, a possible contribution of the confounding effect cannot be ruled out at this stage. Nucleotide metabolism, particularly the tryptophan-NAD⁺ pathway, also seems to be altered in cases of malformations supported by a decrease in 2-Py, along with higher NMND (Figure 6.9). Higher NMND levels were also found in urine of women carrying malformed fetuses (section 4.1.1), suggestive of altered nucleotide metabolism. The decreased 2-Py and increased NMND levels found in this cohort are consistent with a perturbation in nucleotide metabolism in cases of malformations, as previously detected in the prenatal period through maternal urine.

6.4 Impact of maternal disorders on newborn urine composition

The follow-up of pregnancies and their relationship with the newborn health is of paramount importance. Indeed, the importance of *in utero* fetal programming (i.e. the understanding of the permanent effect of the prenatal period in the organism's physiology and metabolism) is being increasingly recognized as reviewed in the literature (Freeman, 2010, Cianfarani *et al.*, 2012). It is known that prenatal disorders affecting maternal-fetal placental circulation, maternal glucose metabolism and infection compromise fetal growth, predisposing infants to increased medical intervention (Hod *et al.*, 2008, Kliegman *et al.*, 2011). In this study, information regarding future health of the newborns of this cohort was not available at this stage. Still, it was possible to assess the impact of GDM and maternal psychiatric disorders in the newborns health status and metabolism during the first days of life.

6.4.1 Maternal Gestational Diabetic Mellitus (GDM)

GDM poses significant risks for the fetus during the prenatal period, including congenital malformations, macrosomia, organomegaly, CNS development delay, chronic hypoxemia, stillbirth, preterm birth and birth injury (Hod *et al.*, 2008). Furthermore, major short-term implications for the newborn are associated with being born to GDM mothers, namely hypoglycemia, hypocalcemia, hypomagnesemia, polycythemia,

hyperbilirubinemia, respiratory distress syndrome (RDS) and vascular thrombosis (Hod *et al.*, 2008). One important mention concerning the designation of this group is that in literature being born to a GDM mother is often called iGDM the prefix *i* standing for infant. As in this study all samples were collected from day 1 to day 6 of life, i.e. within the age range considered newborn (from birth to 1st month (Kliegman *et al.*, 2011)) the term newborn was adopted to describe this group in this thesis.

6.4.1.1 Unveiling an urinary metabolic signature of GDM

In this cohort, urine was collected from newborns born to mothers affected by GDM ($n=14$, out of these only 2 were born LGA) and compared to controls. Visual comparison of the average NMR spectra of cases and controls revealed differences in succinate, creatine and formate contents (noted in Figure 6.1d, page 212).

PCA and PLS-DA was applied to both original and reduced dataset and model validation was achieved by MCCV (model quality parameters compiled in Table A-III.4, Annex III). PLS-DA revealed a separation trend between classes in the corresponding score plot, this not being found in the PCA (Figure 6.19a,b). The PLS-DA model was found to have relatively good CR (82%), sensitivity (73%) and specificity (83%), along with median Q^2 of 0.401 of true models. These parameters, along with the low overlap of Q^2 values obtained for true and permuted models as well as the ROC plot (Figure 6.19c,d) showed the models' relatively satisfactory robustness and predictive power, thus enabling a possible metabolic hypothesis to be advanced.

The urinary metabolic profile of being born to a GDM mother was characterized by changes in 21 metabolites (comprising 9 metabolites derived from diet and/or gut microflora, 3 ketone bodies, 2 amino, 2 organic acids, 2 products of nucleotide metabolism and 3 others) 4 of them having $p<0.05$ (namely 2-KG, ethanolamine, hippurate, IS), 1 unassigned singlet at 2.81 ppm (U15) and the unassigned region of 8.04-8.06 ppm (Table 6.9). An interesting finding is the common variation found for 2-KG (increased) and hippurate (decreased) which were also found in maternal urine of diagnosed GDM women (section 5.1), these being noted with * in Table 6.9.

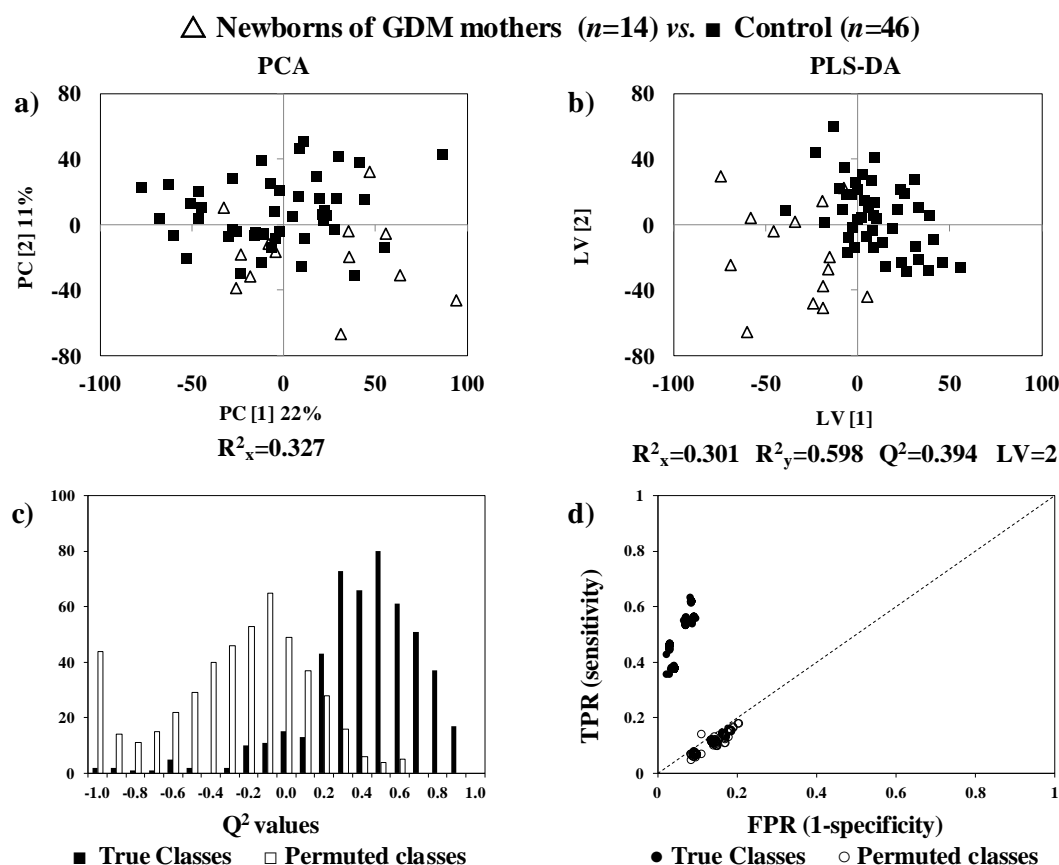


Figure 6.19: a) PCA and b) PLS-DA score plot, c) Q^2 values and d) ROC plot of true and permuted models obtained for newborn urine corresponding to newborns of GDM mothers ($n=14$) compared to controls ($n=46$). TPR: true positive rate, FPR: false negative rate.

The urinary signature of GDM was also plotted by the VIP-wheel and Volcano plots (shown in Figure 6.20a,b) with metabolites/resonances relevant for this profile highlighted. The VIP-wheel shown in Figure 6.20a shows that lysine, acetone, succinate, 2-KG, ethanolamine, 4-HPA, IS, hippurate, U18 and unassigned regions 1.75-1.86 ppm, 6.32-6.37 ppm and 8.26-8.28 ppm have the highest importance for the group separation obtained through PLS-DA. Out of the previous metabolites, only 2-KG, IS, ethanolamine, hippurate, U18 along with the unassigned region 8.04-8.06 ppm were found to have high individual biological and statistical significance (effect size higher than $|0.6|$ and p -values in the order of 10^{-2}) as visible in the Volcano plot shown in Figure 6.20b. An interesting note is that lysine, acetone, 4-HPA and unassigned region 1.75-1.86 ppm, which were found important for the PLS-DA separation were associated with high effect size ($>|0.6|$) but not $p < 0.05$, once more stressing the importance of considering the whole profile rather than metabolites with individual statistically significant.

Table 6.9: Metabolite/resonance changes in newborn urine of newborns from GDM mothers. ^a Chemical shifts selected by variables selection; s: singlet, d: doublet, dd: doublet of doublets, t: triplet, m: multiplet. ^b Peaks possibly related to diet and/or gut microflora. ^{*} Variations found commonly with maternal urine of GDM women. Ui: unassigned compound i by order of appearance and following Table 6.2, Table 6.4, Table 6.5, Table 6.6, Table 6.7 and Table 6.8. Abbreviations defined in Table 6.2, Table 6.4, Table 6.5, Table 6.6, Table 6.7 and Table 6.8. Only p -values<0.05 are shown.

Newborns of GDM mothers (n=14) vs. Control (n=46)		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, p -value)
1-Me-histidine ^b	3.72 s, 7.78 s	↓(-0.16)
2-KG [*]	3.01 t	↑(0.90, 1.6×10^{-2})
3-HBA	2.31 m, 4.15 m	↓(-0.49)
4-HPA ^b	3.46 s, 6.86 d, 7.17 d	↑(0.70)
Acetoacetate	2.28 s	↓(-0.39)
Acetone	2.23 s	↓(-0.63)
Alanine	1.49 d	↑(0.55)
Allantoin	5.39 s	↓(-0.29)
Ascorbate	4.01 m, 4.52 d	↑(0.47)
Creatine	3.04 s, 3.93 s	↑(0.33)
DMA ^b	2.72 s	↓(-0.39)
Ethanolamine	3.15 t, 3.84 t	↓(-0.75, 1.2×10^{-2})
Hippurate ^{b, *}	3.97 d, 7.56 t, 7.64 t, 7.84 d	↓(-0.78, 1.7×10^{-2})
IS ^b	7.21 dd, 7.28 dd, 7.51 d, 7.70 d	↑(0.91, 4.9×10^{-2})
Lactose ^b	4.46 d, 5.25 d	↓(-0.45)
Lysine	1.73 m, 1.92 m	↑(0.62)
NMND	8.78 m, 8.87 d, 8.90 d, 9.28 s	↑(0.51)
PAG ^b	7.36 t	↑(0.37)
Scyllo-inositol ^b	3.36 s	↑(0.35)
Succinate	2.41 s	↓(-0.40)
Xylose ^b	4.59 d, 5.21 d	↓(-0.10)
U15	2.81 s	↑(0.76, 1.4×10^{-2})
Unassigned spectral regions		
8.04-8.06		↑(0.62, 2.0×10^{-2})
+14 spin systems and 12 spectral regions with $p > 0.05$		
Literature review:		
Changes in umbilical cord serum in GDM cases: ↓ glucose, ↑ pyruvate, histidine, alanine, valine, methionine, arginine, lysine, hypoxanthine, lipoprotein and lipids (Dani <i>et al.</i> , 2013).		

6.4.1.2 Proposed metabolic interpretation of changes related with GDM

The set of metabolites with higher biological and statistical relevance enables a potential interpretation of the metabolic status of the newborn born to GDM mothers to be put forward. Significantly higher (p -value<0.05) 2-KG and IS levels were found in newborns of GDM mothers, reflecting respectively changes in TCA and oxidative stress, whilst decreased ethanolamine and hippurate suggest changes in lipids metabolism and gut microflora, respectively. Moreover, significantly higher 2-KG was also found in maternal

urine collected at the time of GDM diagnosis, whilst decreased hippurate was found at diagnosis and at the *pre*-diagnosis states (discussed in Chapter 5, sections 5.1 and 5.2).

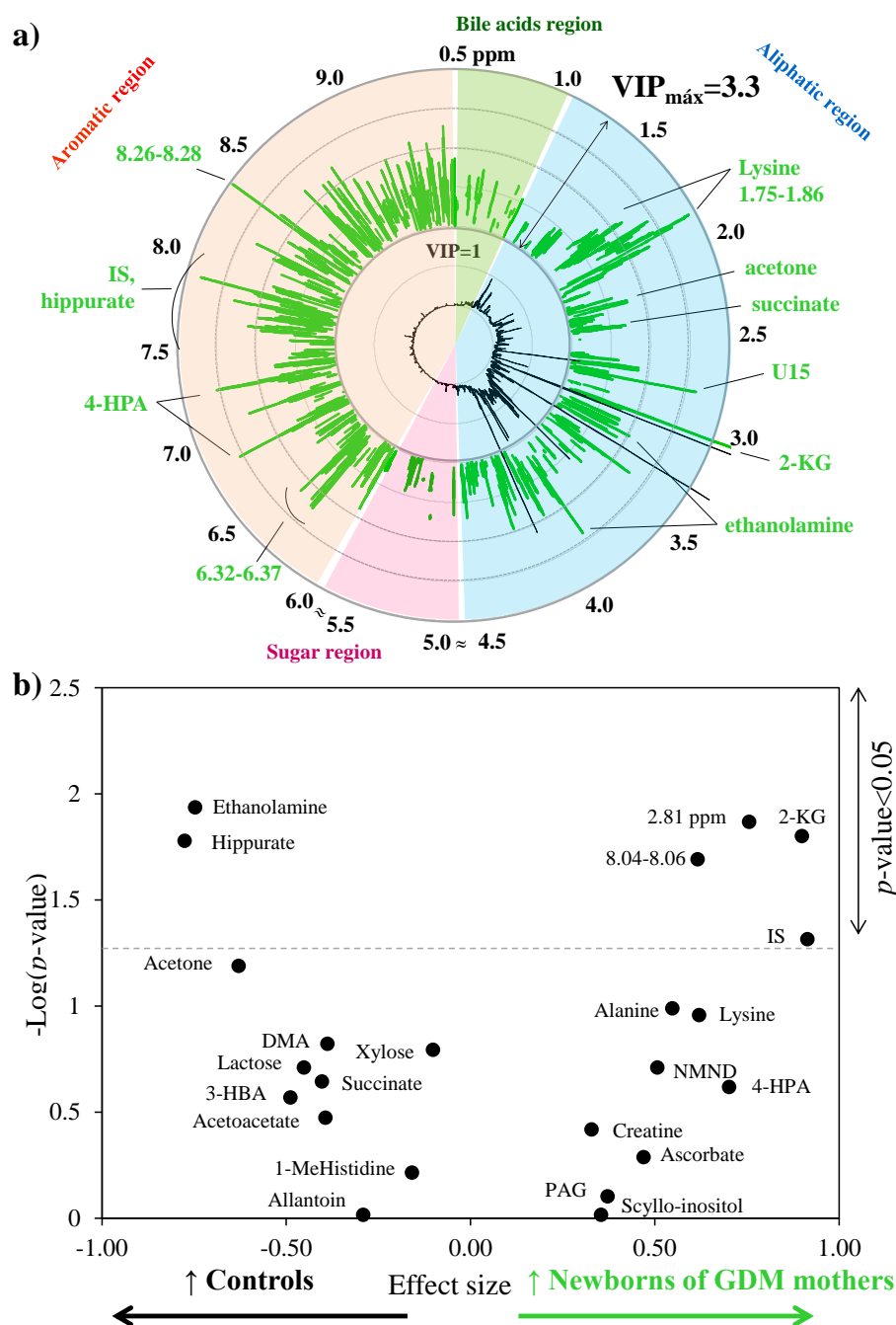


Figure 6.20: a) VIP-wheel representation of the NMR urinary profile of newborns born to mother with GDM (inner and outer circles represent, respectively, the average urine spectrum of healthy newborn urine (control group) and corresponding ppm scale). b) Volcano plot of effect size vs. $-\log(p\text{-value})$ of metabolite/resonances integrals listed in Table 6.9. Dashed line indicates $p\text{-value}=0.05$.

Higher oxidative stress in newborn from GDM-affected pregnancies is here hypothesized by the significantly higher urinary excretion of IS. This suggestion is in agreement with higher oxidative stress, based on higher hypoxanthine levels in umbilical cord blood, previously reported in infants of GDM-affected pregnancies (Dani *et al.*, 2013). Increased excretion of 2-KG in newborn urine suggests an increase in the TCA cycle, as previously proposed for maternal metabolism in cases of GDM (Chapter 5, section 5.1). Additionally, the common variation (significant increase) of 2-KG in maternal and newborn biofluids can be a reflection of a metabolic deviation affecting both mother and child in GDM pregnancies. This common deviation was also verified for hippurate which decreases significantly in maternal (during pregnancy) and newborn urine in relation to GDM. Lower hippurate in maternal urine was found in GDM cases as aforementioned in Chapter 5. Hippurate is a metabolite derived from the gut microflora, the latter known to play a relevant role in diabetes. Thus, the significant decrease of urinary hippurate here found suggests an involvement of the fetal/newborn gut microflora, possibly reflecting a different microbial colonization, as a consequence of altered maternal microflora.

Decreased ethanolamine was found in this cohort, this metabolite being a precursor of phosphatidylethanolamine, a membrane phospholipid also acting as lung surfactant. Lower ethanolamine was also found in the premature group possibly reflective of altered surfactant synthesis. However, altered lung surfactant production is an unlikely source for the ethanolamine decrease found in newborns of GDM mothers, as only 1 case was born premature and 1 other case was affected by a respiratory depression episode after birth. However, in relation to GDM, it has been previously reported that umbilical cord serum of newborns from GDM mothers have higher contents of lipoproteins and lipids as a result of altered fatty acid metabolism in the placenta or insulin action (Dani *et al.*, 2013). Thus, it seems possible that the ethanolamine decrease found in this cohort could be a reflection of altered lipid metabolism, particularly decreased phosphatidylethanolamine synthesis, rather than perturbations in lung surfactant production.

6.4.2 Maternal psychiatric disorders

The number of samples available for the remaining groups of maternal disorders were very low ($n \leq 10$, Experimental Section). Still, given the importance of the prenatal

period in predisposing future health, i.e. fetal programming, an exploratory study of the impact of maternal respiratory, thyroid, hypertension and psychiatric disorders in newborns health status was attempted. PCA and PLS-DA models were computed and validated using the original datasets and after variable selection, the best result being always achieved with the compressed matrix (see Table A-III.4, Annex III). Unfortunately, the detailed interpretation of variables subsets and the unveiling of the urinary metabolic signatures for each of these maternal disorders were not possible in time for this thesis, this subject being surely accomplished in a near future. Yet, an initial exploitation of the effect of maternal psychiatric disorder in newborns health was attempted. This preliminary analysis was only possible for the group of maternal psychiatric disorders, this having been chosen for its clinical relevance and to check if valuable information could be retrieved even in cases of such limited number of samples.

6.4.2.1 Unveiling an urinary metabolic signature of maternal psychiatric disorders

Urine of newborns born to mothers with psychiatric disorders (comprising depression and epilepsy) was compared to controls through PCA and PLS-DA only the latter showing a separation trend between classes (score plots being shown Figure 6.21a,b). The corresponding MCCV results are shown in Figure 6.21c,d, these showing the models' predictive power (Q^2 distribution and median Q^2 0.549) and robustness with CR 91%, sensitivity 96% and specificity 90%. Regarding the distribution of true models in the ROC plot (noted in Figure 6.21d), the abnormal scattering seen through the left limit (FPR=0 and TPR approximately from 0.2-0.6) is an artefact caused by the significant unbalance of the groups (46 vs. 7 samples).

The suggested urinary metabolic signature is shown Table 6.10. Overall, 15 metabolites, comprising 2 products of nucleotide metabolism, 3 organic acids, 3 amino acids and derivatives, 1 ketone body, 4 metabolites deriving from diet and/or gut microflora and 2 others, were found to characterize this group along with 1 unassigned singlet at 2.87 ppm and 2 unassigned regions with $p < 0.05$. Metabolites/resonances with

higher VIP, biological and statistical relevance are highlighted in the VIP-wheel and volcano plot shown in Figure 6.22a,b.

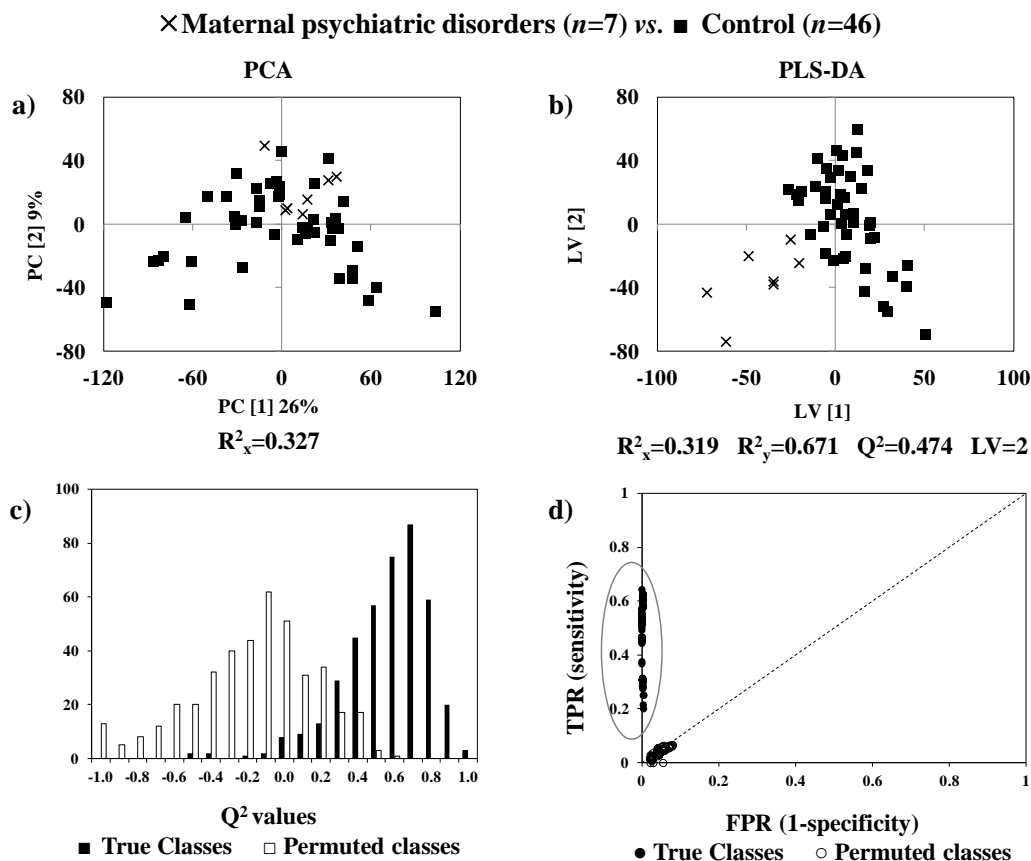


Figure 6.21: a) PCA and b) PLS-DA score plot, c) Q^2 values and d) ROC plot of true and permuted models obtained for newborn urine corresponding to the maternal psychiatric disorder group ($n=7$) compared to controls ($n=46$). TPR: true positive rate, FPR: false negative rate.

The VIP-wheel (Figure 6.22a) revealed that metabolites/resonances with highest importance for the PLS-DA group separation were DMG, 2-KG, *cis*-aconitate, xylose, U19 and unassigned regions 1.78-1.92, 3.37-3.39, 4.15-4.18, 8.01-8.05 and 8.26-8.30 ppm. Within the previous list only DMG, 2-KG, U19, as well as unassigned regions 1.88- 1.92 and 8.26-8.30 were found to have high biological (effect size $>|0.7|$) and statistical significance (p -value < 0.05). Other relevant metabolites include *cis*-aconitate, hippurate and taurine, which were found to have high biological but not statistical significance (effect size $>|0.5|$, p -value > 0.05), as visible through the Volcano plot (Figure 6.22b).

Table 6.10: Metabolite/resonance changes in newborn urine of newborns from mother with psychiatric disorders.^a Chemical shifts selected by variables selection; s: singlet, d: doublet, dd: doublet of doublets, t: triplet, m: multiplet, br: broad. ^b Peaks possibly related to diet and/or gut microflora. Ui: unassigned compound i by order of appearance and following Table 6.2, Table 6.4, Table 6.5, Table 6.6, Table 6.7, Table 6.8 and Table 6.9. Abbreviations defined in Table 6.2, Table 6.4, Table 6.5, Table 6.6, Table 6.7, Table 6.8 and Table 6.9. Only p -values<0.05 are shown.

Maternal psychiatric disorders (n=7) vs. Controls (n=46)		
Compound	δ_H /ppm and multiplicity ^a	Variation (effect size, p -value)
2-KG	2.45 t, 3.01 t	↓(-0.74, 4.6×10^{-2})
2-Py	8.33 d, 7.97 dd	↑(0.52)
3-aminoisobutyrate	1.19 d	↑(0.23)
3-HBA	1.20 d, 2.31 m, 2.41 m, 4.15 m	↓(-0.31)
4-DEA	1.11 d	↓(-0.23)
4-HPA ^b	6.86 d, 7.17 d	↓(-0.42)
Adipate	1.55 m	↓(-0.16)
Cis-aconitate	3.12 d	↑(0.63)
Creatine	3.04 s, 3.93 s	↑(0.53)
DMG	2.93 s, 3.72 s	↑(1.50, 2.6×10^{-2})
Ethanolamine	3.84 t	↓(-0.48)
Galactose ^b	5.28 d	↓(-0.48)
Hippurate ^b	7.55 t, 7.64 t, 7.84 d, 8.52 br	↓(-0.58)
Lactose ^b	4.46 d	↓(-0.48)
Taurine	3.43 t	↓(-0.56)
U19	2.87 s	↑(0.80, 2.6×10^{-2})
Unassigned spectral regions		
1.88-1.92		↑(0.74, 4.0×10^{-2})
8.26-8.30		↑(0.89, 3.5×10^{-2})
+13 unassigned spin systems and 14 spectral regions with $p > 0.05$		

6.4.2.2 Proposed metabolic interpretation of the changes related with maternal psychiatric disorders

Within the registered changes, DMG was found to significantly increased (effect size 1.50, p -value<0.05) whilst 2-KG was found decreased in the pathological group. The increase noted for DMG in tandem with decreased taurine suggest a perturbation of the methionine cycle, as DMG is formed through the methylation of homocysteine whilst its transsulfuration forms taurine. Also, the decrease of 2-KG and increased *cis*-aconitate are consistent with changes in TCA cycle activity. Unfortunately, given the exploratory nature of the analysis of this group, no further hypothesis can be advance at this stage. However, these results suggest that maternal psychiatric disorders affecting the mother during the pregnancy pose a metabolic effect on the newborn, which needs to be further explored using larger cohorts.

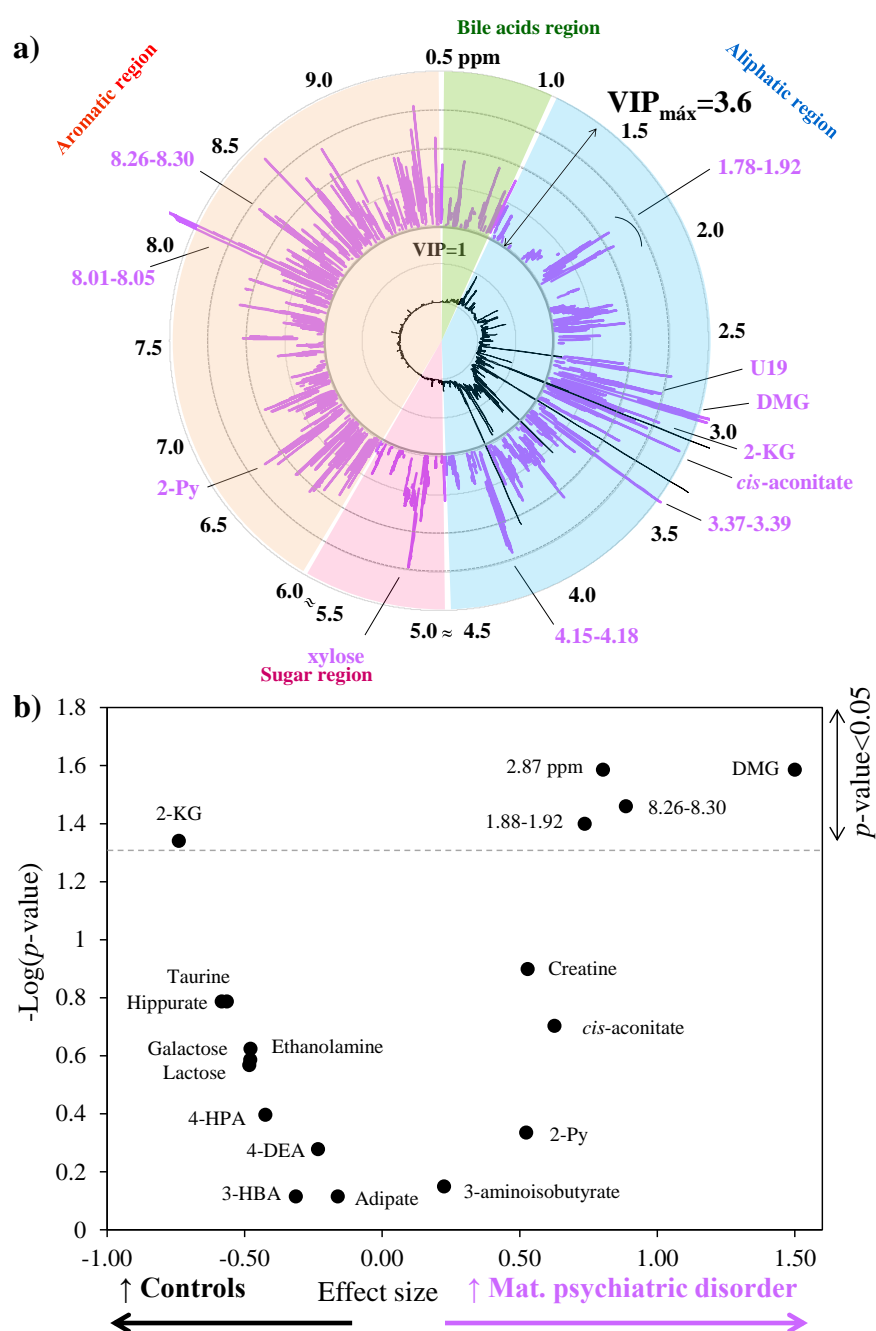


Figure 6.22: a) VIP-wheel representation of the NMR urinary profile of newborns from mother with psychiatric disorders (inner and outer circles represent, respectively, the average urine spectrum of healthy newborn urine (control group) and ppm scale). Main differences are noted. b) Volcano plot of effect size vs. $-\log(p\text{-value})$ of metabolite/resonances integrals shown in Table 6.10 (dashed line indicates $p=0.05$ defined for statistical significance). Mat: maternal.

6.5 Summary of newborns and maternal disorders in newborn urine

The results presented in section 6.2 showed that the confounding effect with highest impact in newborn urine was delivery mode followed by gender, these effects being necessarily taken into account when variation in pathological groups (unbalanced in terms of delivery mode and gender) were interpreted. Regarding newborns' and maternal disorders (section 6.3 and 6.4), prematurity was found to pose the highest impact in newborn urine composition followed by respiratory depression, LGA, newborns from GDM mothers, newborns carrying malformations and maternal psychiatric disorder (the latter being a preliminary and exploratory study due to limited number of samples available). The main biochemical findings for these groups along with the validation parameters are summarized and presented in decreasing order of disorder impact in Figure 6.23. An interesting result was that similar variations were found in both mothers (maternal prenatal urine) and newborns (urine collected at birth), in cases of prematurity (increased citrate and 4-OH-hippurate, decreased 3-Me-histidine and creatinine), babies carrying malformations (increased NMND, decreased succinate) and newborns born to GDM-affected pregnancies (increased 2-KG, decreased hippurate). This indicates that these conditions pose specific metabolic perturbations that affect both mother and child, which may be detected in the prenatal period (through maternal urine) and confirmed after birth in newborn urine.

The abovementioned figure shows that the highest impact on newborn urine was found for prematures and respiratory depression groups, with CR, sensitivity and specificity above 90%. Then, LGA and GDM were found to pose the following higher impact, with CR and specificity 82-86%, whilst sensitivity ranged from 73-76%. This lower sensitivity expresses a low true positive rate (TPR), meaning that true cases (LGA or GDM cases) are sometimes misclassified as controls. The effect of carrying a malformation was associated with high CR and sensitivity (high TPR) although with lower specificity (73%). A low specificity reflects a high false positive rate (FPR), meaning that some control samples are misclassified as malformations.

Disorder studied	Main biochemical characteristics	Validation parameters
Prematurity	Nucleotide and amino acids metabolism, TCA cycle, renal function, oxidative stress and perturbations in lung surfactant production	CR: 92% Sensitivity: 92% Specificity: 92%
Respiratory depression	Methionine metabolism, altered lung surfactant production and/or altered myelination	CR: 91% Sensitivity: 96% Specificity: 91%
LGA	Methionine cycle, nucleotide and bile acids metabolism	CR: 84% Sensitivity: 76% Specificity: 86%
GDM	Oxidative stress, TCA cycle, lipids metabolism and gut microflora	CR: 82% Sensitivity: 73% Specificity: 83%
Malformations	Respiratory chain, TCA cycle and nucleotide metabolism	CR: 89% Sensitivity: 90% Specificity: 73%
Maternal psychiatric disorders	TCA cycle and methionine metabolism	CR: 91% Sensitivity: 96% Specificity: 90%

Figure 6.23: Schematic summary of the disorders studied, main biochemical findings and validation parameters, listed in decreasing order of impact. CR: classification rate.

A schematic representation of the main metabolic pathways proposed to be affected in relation to delivery mode, gender, and newborns' and maternal disorders, in newborn urine composition, is shown in Figure 6.24. In this figure, colored arrows indicate metabolites variations for each condition here studied. The concomitant analysis of all disorders here studied revealed specific metabolite changes for selected conditions, namely increased glucose in females, decreased acetoacetate in newborns from GDM mothers, increased glutamine in babies carrying malformations, decreased hypoxanthine after a respiratory depression episode and decreased lactate and creatinine in premature newborns. Another notable feature in Figure 6.24 is the decrease of 3-HBA in all disorders (while independent of delivery mode and gender), this possibly reflecting a general effect of a pathological situation.

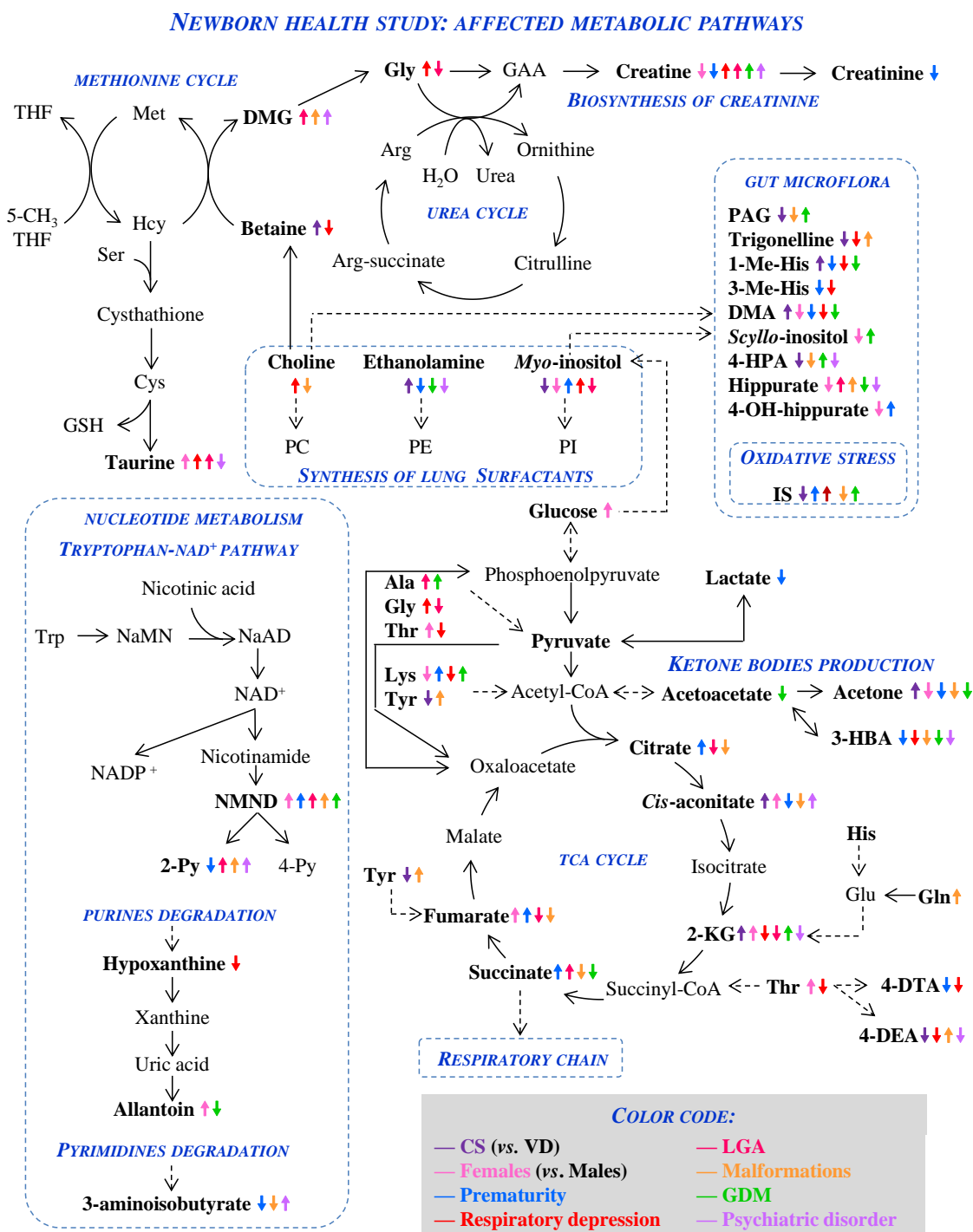


Figure 6.24: Schematic representation of the main metabolic pathways affected in cases of newborn and maternal disorders, seen through newborn urine. Amino acids in 3-letter code; NaMN: nicotinic acid mono nucleotide, NaAD: nicotinic acid adenine dinucleotide, NAD⁺: nicotinamide adenine dinucleotide, NADP⁺: nicotinamide dinucleotide phosphate, ADP: adenosine diphosphate, NMND: *N*-methyl-nicotinamide, 2-Py: *N*-methyl-2-pyridone-5-carboxamide, 4-Py: *N*-methyl-4-pyridone-3-carboxamide, 4-DEA: 4-deoxyerythronic acid, 4-DTA: 4-deoxythreonic acid, 3-HBA: 3-hydroxybutyrate, PAG: phenylacetylglutamine, DMA: dimethylamine, PI: phosphatidylinositol, PE: phosphatidylethanolamine, PC: phosphatidylcholine, DMG: dimethylglycine, GAA: guanidoacetate, Hcy: homocysteine, GSH: glutathione, THF: tetrahydrofolate.

7. Conclusions and future perspectives

This work employed a metabolomics approach to characterize maternal and newborn urine composition of healthy and pathological cases, mainly using NMR spectroscopy but also MS in selected instances. The results obtained are presented in Chapters 3 to 6, corresponding to a set of four interconnected studies.

The first study aimed at defining and characterizing the urinary metabolic profile accompanying healthy pregnancies, to deepen the understanding of the dynamic metabolic adaptations taking place during healthy pregnancies. The results were achieved by comparing independent groups of healthy subjects (representing the non-pregnant state and each of the three pregnancy trimesters) and a 21-metabolite signature characteristic of each trimester was unveiled. Within the variations noted, those affecting choline, creatinine, 4-deoxyerythronic acid (4-DEA), 4-deoxythreonic acid (4-DTA), furoylglycine, guanidoacetate (GAA), 3-hydroxybutyrate (3-HBA) and lactate were observed here for the first time, to our knowledge, in connection to pregnancy. Other previously known effects were confirmed, namely the tendency for aminoaciduria and changes in lipid and methionine metabolisms. This work enabled the definition of a dynamic metabolic signature describing healthy pregnancies, unveiling novel metabolic aspects and setting the basis for a control trajectory against which disease-related deviations may be confronted, serving as the basis for improved disease diagnostics and prediction.

All results described below, entailing the analysis of maternal urine in relation to pregnancy disorders, are novel in comparison to literature, with the exception of one previous report on gestational diabetes. Hence, to our knowledge, this is the first work to explore and demonstrate the value of maternal urine as a means to diagnose and/or predict prenatal disorders.

NMR metabolomics of 2nd trimester maternal urine enabled metabolic signatures to be unveiled for diagnosed FM (and, specifically of central nervous system (CNS) malformations) and CD (and, specifically, of trisomy 21). For FM cases, some of the results found were in agreement with previously advanced hypotheses of enhanced gluconeogenesis under hypoxia, enhanced TCA cycle operation and underuse of the respiratory chain, along with increased lipid oxidation and ketone body production. A

novel observation involved changes noted in nucleotide metabolism, specifically the tryptophan-NAD⁺ pathway, suggested by the increased excretion of *N*-methylnicotinamide (NMND). CNS malformations were related to known perturbations in methionine metabolism (with connection to the folic acid pool) and in the respiratory chain (through changes in succinate). As previously suggested, changes in amino acid metabolism was here found to translate into a profound deregulation, based on higher excretion of branched chain amino acids (BCAA, leucine, isoleucine and valine), alanine and glutamine (only glutamine and valine having been previously found to vary in relation to CNS). Furthermore, new evidence of enhanced TCA cycle (namely decreased acetate along with increased 2-ketoglutarate (2-KG) and citrate) and increased oxidative stress (higher excretion of allantoin and indoxyl sulphate-IS) were found in CNS cases. CD cases were related to enhanced ketone body production, derangements in sugars and amino acid metabolism, consistent with previous knowledge. In addition, new perturbations were registered in connection to nucleotide metabolism (supported by higher NMND excretion), TCA cycle operation (increased citrate and 2-KG), urea cycle and/or creatine biosynthesis (higher excretion of GAA, creatine and creatinine). The urinary profile found for T21 cases reflected known perturbations in amino acid metabolism, especially for taurine (related with the methionine cycle) and allantoin (reflecting increased oxidative stress). Additionally, perturbations in respiratory chain (decreased malonate) and nucleotide metabolism (decreased *N*-methyl-2-pyridone-5-carboxamide (2-Py) and NMND) were newly found.

Bearing added potential clinical interest, predictive signatures were found in 2nd trimester maternal urine for *later developing* preterm delivery (PTD), intrauterine growth restriction (IUGR) and preeclampsia (PE), whilst no early evidence of premature rupture of the membranes (PROM) was found. Changes in *pre*-PTD urine confirmed known perturbations in amino acid metabolism in the 2nd trimester and/or placental transfer and indications of lipids oxidation and oxidative stress previously noted at delivery. Regarding *pre*-IUGR and *pre*-PE profiles, some novel aspects of the conditions were advanced, namely decreased amino acids in *pre*-IUGR samples and early perturbations in TCA cycle in *pre*-PE cases.

All disease studies were carried out by considering subsets of ¹H NMR resulting from a novel variable selection method developed in the context of this thesis, this being

one of the major outcomes of this work. Regarding validation procedures, internal validation was performed for all models and external validation was attempted in the cases of FM and *pre*-PTD (for which higher sample numbers were available). The latter did not, however, yield reliable prediction results within the duration of this thesis, probably due to a limitation posed by the necessary pre-processing step of spectral alignment, alternative approaches having been discussed. For *pre*-PTD, an additional study was carried out using UPLC-MS, based on the premises that enhanced sensitivity (sub mM to pM) might bring new information on the disease. MS data processing for multivariate analysis was found to be an important issue, different data processing methodologies having been tested. The best performing method consisted of the application of specific filters to select reliable data, however, no significant additional information was obtained, only a possible biomarker compound being suggested by statistical correlation (STOCSY) results.

A third study focused on the impact of Gestational Diabetes Mellitus (GDM) on maternal urine, comprising the comparison of samples collected at diagnosis (16-33 g.w.), prior to diagnosis or *pre*-diagnosis (15-22 g.w.) and upon treatment (17-40 g.w.), compared to gestational age-matched controls. This study provided a better understanding of maternal metabolism in GDM pregnancies, showing changes in glucose and energy metabolisms, renal function and gut microflora, also found in type 1 and type 2 diabetes. The metabolic response to GDM treatment was studied and results showed that maternal metabolism does not completely return to the normal situation, with accountable changes in TCA cycle, amino acids metabolism and gut microflora. Moreover, 2nd trimester maternal urine was found to carry information about metabolic changes *preceding* the clinical manifestation of GDM, supporting a possible application of urine metabolomics in the prediction of this disease. These changes confirmed an *pre*-diagnosis hyperglycemia tendency and increased urinary citrate, and unveiled new changes in ketone bodies excretion, fatty acids β -oxidation, amino acids and nucleotide metabolism, altered biotin status and perturbation in choline and some gut microflora metabolites. The possibility of predicting individual GDM treatment requirements (based on the pharmacometabolomics concept), was explored for the first time and, indeed, a *pre*-treatment metabolic phenotype was found in maternal urine. This suggested that women later requiring insulin therapy exhibit a *pre-treatment* signature comprising evidence of TCA slowing down and ketone

bodies production, decreased malonate (in connection with respiratory chain), changes in nucleotide metabolism (higher 2-Py and NMND), higher creatine and creatinine, and metabolites related with gut microflora. This *pre-treatment* signature can be helpful in predicting individual response to future treatment requirements, thus allowing an earlier and more effective intervention.

The impact of pregnancy and newborn characteristics on newborn urine composition was presented next. In this work, phenylacetylglutamine (PAG) and indoxyl sulphate (IS) were identified in newborn urine for the first time to our knowledge. A significant impact was found in relation to delivery mode and gender, for the first time to our knowledge. The changes observed are in agreement with the known higher oxidative stress and changes in gut microflora colonization in babies born through vaginal delivery, along with decreased hepatic gluconeogenesis in babies born through cesarean section. Regarding newborn's gender, only acetone, dimethylamine (DMA) and fumarate varied consistently with changes reported in adult cohorts. In addition, female newborns were found to have higher urinary allantoin and *myo*-inositol was higher in males, indicating novel gender-related differences in sugar metabolism and oxidative stress. Regarding the effect of the *in utero* period and subsequent newborn disorders, respiratory depression after birth, LGA and babies with malformations were here studied for the first time to our knowledge, along with prematurity and GDM (the latter previously studied based on umbilical blood only). Prematurity was found to cause the highest impact on urine composition, with changes in nucleotides and purine metabolism, lipids β -oxidation, renal function, oxidative stress and new evidence of perturbations in pyrimidine degradation and lung surfactant production. Subtler changes were also noted in the urine of newborns affected by respiratory depression and LGA. Interestingly, for premature newborns, FM babies and those born to GDM-affected mothers, similar changes were noted in the urine of newborns and mothers during pregnancy. The last group to be studied (still of a limited size, $n=7$) was that of newborns from mothers with psychiatric disorders, which surprisingly suggested changes in 2-KG and DMG, involved in the TCA and methionine cycle, respectively. These studies will be pursued in the future, as will those of other maternal disorders such as respiratory, thyroid and hypertension (only initially explored in this thesis).

An important general conclusion of this work was the recognition of the paramount importance of considering the full spectral profiles provided by NMR, which correspond to disease fingerprints of great complexity, instead of a limited number of discrete signals (or metabolites) with individual statistical relevance ($p < 0.05$). Such full spectral profile almost invariably includes regions or signals of low intensity and difficult assignment, yet playing a determinant role in model robustness and sample classification.

The work presented in this thesis demonstrates the promising use of urine metabolomics for detection and prediction of disorders of the mother, fetus and newborn. Due to its non-invasiveness and ease of collection, maternal and newborn urine emerges as an invaluable matrix to search for novel markers with potential clinical applications. It is sincerely hoped that the work described in this thesis can indeed contribute towards the improvement of prenatal and newborns healthcare. With this in mind, future work will necessarily entail the use of larger cohorts to confirm the changes found in this work for 2nd trimester urine, specially for T21, and for the prediction of PE, PTD, and fetal growth disorders (IUGR, LGA and also small for gestational age). Also, larger sample numbers will be required to confirm the enticing possibility of predicting GDM development as well as individual GDM treatment requirements suggested in this work. The use of larger cohorts will necessarily entail appropriate external validation so that a clinical application can be envisaged. Another future endeavor is to search for markers of these disorders in 1st trimester maternal urine, thus enabling earlier diagnosis and intervention. Finally, regarding infant health, the implementation of follow-up strategies for regular collection of subject clinical data from day 1 of life onwards (1-2 first years of life or later) would be of great value for the prediction of shorter- (1-2 years of life) and longer-term effects of pregnancy characteristics.

8. Bibliography

- Agassandian, M. and Mallampalli, R.K., Surfactant phospholipid metabolism, *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, **2013**, 1831(3): 612-625.
- American Diabetes Association (ADA), Gestational Diabetes Mellitus, *Diabetes Care*, **2004**, 27(suppl 1): s88-s90.
- American Diabetes Association (ADA), Standards of medical care in diabetes- 2013, *Diabetes Care*, **2013**, Suppl 1S11-66.
- Alexandre-Gouabau, M.-C., Courant, F., Moyon, T., Küster, A., Le Gall, G., Tea, I., Antignac, J.-P. and Darmaun, D., Maternal and cord blood LC-HRMS metabolomics reveal alterations in energy and polyamine metabolism, and oxidative stress in very-low birth weight infants, *Journal of Proteome Research*, **2013**, 12(6): 2764-2778.
- Alm, E., Torgrip, R.J.O., Aberg, K.M., Schuppe-Koistinen, I. and Lindberg, J., A solution to the 1D NMR alignment problem using an extended generalized fuzzy Hough transform and mode support, *Analytical and Bioanalytical Chemistry*, **2009**, 395(1): 213-223.
- Amaral, A.U., Cecatto, C., Busanello, E.N.B., Ribeiro, C.A.J., Melo, D.R., Leipnitz, G., Castilho, R.F. and Wajner, M., Ethylmalonic acid impairs brain mitochondrial succinate and malate transport, *Molecular Genetics and Metabolism*, **2012**, 105(1): 84-90.
- Amorini, A., Giorlandino, C., Longo, S., D'Urso, S., Mesoraca, A., Santoro, M., Picardi, M., Gullotta, S., Cignini, P., Lazzarino, D., Lazzarino, G. and Tavazzi, B., Metabolic profile of amniotic fluid as a biochemical tool to screen for inborn errors of metabolism and fetal anomalies, *Molecular and Cellular Biochemistry*, **2012**, 359(1): 205-216.
- Andersen, C.M. and Bro, R., Variable selection in regression—a tutorial, *Journal of Chemometrics*, **2010**, 24(11-12): 728-737.
- Appiah-Amponsah, E., Shanaiah, N., Nagana Gowda, G.A., Owusu-Sarfo, K., Ye, T. and Raftery, D., Identification of 4-deoxythreonic acid present in human urine using HPLC and NMR techniques, *Journal of Pharmaceutical and Biomedical Analysis*, **2009**, 50(5): 878-885.
- Athanasiadis, A.P., Michaelidou, A.-M., Fotiou, M., Menexes, G., Theodoridis, T.D., Ganidou, M., Tzeveleakis, B., Assimakopoulos, E. and Tarlatzis, B.C., Correlation of 2nd trimester amniotic fluid amino acid profile with gestational age and estimated fetal weight, *Journal of Maternal-Fetal and Neonatal Medicine*, **2011**, 24(8): 1033-1038.
- Atzori, L., Antonucci, R., Barberini, L., Locci, E., Marincola, F.C., Scano, P., Cortesi, P., Agostiniani, R., Defraia, R., Weljie, A., Gazzolo, D., Lai, A. and Fanos, V., ¹H NMR-based metabolomic analysis of urine from preterm and term neonates, *Frontiers in bioscience (Elite edition)*, **2011**, 31005-1012.
- Auray-Blais, C., Raiche, E., Gagnon, R., Berthiaume, M. and Pasquier, J.-C., Metabolomics and preterm birth: What biomarkers in cervicovaginal secretions are predictive of high-risk pregnant women?, *International Journal of Mass Spectrometry*, **2011**, 307(1-3): 33-38.
- Bacino, C.A. (2012). "Approach to congenital malformations." Accessed 22 January 2013, from www.uptodate.com
-

- Bahado-Singh, R.O., Akolekar, R., Chelliah, A., Mandal, R., Dong, E., Kruger, M., Wishart, D.S. and Nicolaides, K., Metabolomic analysis for first-trimester trisomy 18 detection, *American Journal of Obstetrics and Gynecology*, **2013**, 209(1): 65.e61-65.e69.
- Bahado-Singh, R.O., Akolekar, R., Mandal, R., Dong, E., Xia, J., Kruger, M., Wishart, D.S. and Nicolaides, K., First trimester metabolomic detection of late-onset preeclampsia, *American Journal of Obstetrics and Gynecology*, **2012**, 208(1): 58.e51-58.57.
- Bahado-Singh, R.O., Akolekar, R., Mandal, R., Dong, E., Xia, J., Kruger, M., Wishart, D.S. and Nicolaides, K., Metabolomics and first-trimester prediction of early-onset preeclampsia, *Journal of Maternal-Fetal & Neonatal Medicine*, **2012**, 25(10): 1840-1847.
- Bahado-Singh, R.O., Akolekar, R., Mandal, R., Dong, E., Xia, J., Kruger, M., Wishart, D.S. and Nicolaides, K., Metabolomic analysis for first-trimester Down syndrome prediction, *American Journal of Obstetrics and Gynecology*, **2013**, 208(5): 371.e371-371.e378.
- Bax, A. and Davis, D.G., MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy, *Journal of Magnetic Resonance (1969)*, **1985**, 65(2): 355-360.
- Bernini, P., Bertini, I., Calabro, A., la Marca, G., Lami, G., Luchinat, C., Renzi, D. and Tenori, L., Are Patients with Potential Celiac Disease Really Potential? The Answer of Metabonomics, *Journal of Proteome Research*, **2011**, 10(2): 714-721.
- Bertram, H.C., Hoppe, C., Petersen, B.O., Duus, J.O., Molgaard, C. and Michaelsen, K.F., An NMR-based metabonomic investigation on effects of milk and meat protein diets given to 8-year-old boys, *British Journal of Nutrition*, **2007**, 97(4): 758-763.
- Bock, J.L., Metabolic profiling of amniotic fluid by proton nuclear magnetic resonance spectroscopy: correlation with fetal maturation and other clinical variables, *Clinical Chemistry*, **1994**, 40(1): 56-61.
- Bonvallot, N., Tremblay-Franco, M., Chevrier, C., Canlet, C., Warembourg, C., Cravedi, J.-P. and Cordier, S., Metabolomics tools for describing complex pesticide exposure in pregnant women in Brittany (France), *Plos One*, **2013**, 8(5): e64433.
- Bouatra, S., Aziat, F., Mandal, R., Guo, A.C., Wilson, M.R., Knox, C., Bjorndahl, T.C., Krishnamurthy, R., Saleem, F., Liu, P., Dame, Z.T., Poelzer, J., Huynh, J., Yallou, F.S., Psychogios, N., Dong, E., Bogumil, R., Roehring, C. and Wishart, D.S., The human urine metabolome, *Plos One*, **2013**, 8(9): e73076.
- Boutsikou, T. and Malamitsi-Puchner, A., Caesarean section: impact on mother and child, *Acta Paediatrica*, **2011**, 100(12): 1518-1522.
- Brosnan, J.T. and Brosnan, M.E., Creatine metabolism and the urea cycle, *Molecular Genetics and Metabolism*, **2010**, 100(Supplement): S49-S52.
- Cai, H.-L., Li, H.-D., Yan, X.-Z., Sun, B., Zhang, Q., Yan, M., Zhang, W.-Y., Jiang, P., Zhu, R.-H., Liu, Y.-P., Fang, P.-F., Xu, P., Yuan, H.-Y., Zhang, X.-H., Hu, L., Yang, W. and Ye, H.-S., Metabolomic analysis of biochemical changes in the plasma and urine of first-episode neuroleptic-naïve schizophrenia patients after treatment with risperidone, *Journal of Proteome Research*, **2012**, 11(8): 4338-4350.
- Calhoun, K.C., Barnhart, K.T., Elovitz, M.A. and Srinivas, S.K., Evaluating the association between assisted conception and the severity of preeclampsia, *ISRN Obstetrics and Gynecology*, **2011**, 2011(ID 928592): 5.
-

- Cardwell, C.R., Stene, L.C., Joner, G., Cinek, O., Svensson, J., Goldacre, M.J., Parslow, R.C., Pozzilli, P., Brigis, G., Stoyanov, D., Urbonaitė, B., Šipetić, S., Schober, E., Ionescu-Tirgoviste, C., Devoti, G., Beaufort, C.E., Buschard, K. and Patterson, C.C., Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies, *Diabetologia*, **2008**, 51(5): 726-735.
- Caussé, E., Pradelles, A., Dirat, B., Negre-Salvayre, A., Salvayre, R. and Couderc, F., Simultaneous determination of allantoin, hypoxanthine, xanthine, and uric acid in serum/plasma by CE, *Electrophoresis*, **2007**, 28(3): 381-387.
- Cavill, R., Keun, H.C., Holmes, E., Lindon, J.C., Nicholson, J.K. and Ebbels, T.M.D., Genetic algorithms for simultaneous variable and sample selection in metabonomics, *Bioinformatics*, **2009**, 25(1): 112-118.
- Cetin, I. and Alvino, G., Intrauterine growth restriction: implications for placental metabolism and transport. A review, *Placenta*, **2009**, 30(Supplement): 77-82.
- Cetin, I., de Santis, M.S.N., Taricco, E., Radaelli, T., Teng, C., Ronzoni, S., Spada, E., Milani, S. and Pardi, G., Maternal and fetal amino acid concentrations in normal pregnancies and in pregnancies with gestational diabetes mellitus, *American Journal of Obstetrics and Gynecology*, **2005**, 192(2): 610-617.
- Chen, J., Zhao, X., Fritsche, J., Yin, P., Schmitt-Kopplin, P., Wang, W., Lu, X., Haring, H.U., Schleicher, E.D., Lehmann, R. and Xu, G., Practical approach for the identification and isomer elucidation of biomarkers detected in a metabonomic study for the discovery of individuals at risk for diabetes by integrating the chromatographic and mass spectrometric information, *Analytical Chemistry*, **2008**, 80(4): 1280-1289.
- Cho, S.-W. and Cha, Y.-S., Pregnancy increases urinary loss of carnitine and reduces plasma carnitine in Korean women., *British Journal of Nutrition*, **2005**, 93:685-691.
- Chu, C.Y., Xiao, X., Zhou, X.G., Lau, T.K., Rogers, M.S., Fok, T.F., Law, L.K., Pang, C.P. and Wang, C.C., Metabolomic and bioinformatic analyses in asphyxiated neonates, *Clinical Biochemistry*, **2006**, 39(3): 203-209.
- Cianfarani, S., Agostoni, C., Bedogni, G., Berni Canani, R., Brambilla, P., Nobili, V. and Pietrobelli, A., Effect of intrauterine growth retardation on liver and long-term metabolic risk, *International Journal of Obesity*, **2012**, 36(10): 1270-1277.
- Claridge, T.D.W., High-resolution NMR techniques in organic chemistry, **1999**, Oxford, Elsevier.
- Clayton-Smith, J. and Laan, L., Angelman syndrome: a review of the clinical and genetic aspects, *Journal of Medical Genetics*, **2003**, 40(2): 87-95.
- Clayton, T.A., Baker, D., Lindon, J.C., Everett, J.R. and Nicholson, J.K., Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism, *Proceedings of the National Academy of Sciences of the United States of America*, **2009**, 106(34): 14728-14733.
- Cloarec, O., Dumas, M.E., Craig, A., Barton, R.H., Trygg, J., Hudson, J., Blancher, C., Gauguier, D., Lindon, J.C., Holmes, E. and Nicholson, J., Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic ^1H NMR data sets, *Analytical Chemistry*, **2005**, 77(5): 1282-1289.
- Coen, M., Holmes, E., Lindon, J.C. and Nicholson, J.K., NMR-based metabolic profiling and metabonomic approaches to problems in molecular toxicology, *Chemical Research in Toxicology*, **2008**, 21(1): 9-27.

-
- Coen, M., Hong, Y.S., Cloarec, O., Rhode, C.M., Reily, M.D., Robertson, D.G., Holmes, E., Lindon, J.C. and Nicholson, J.K., Heteronuclear ^1H - ^{31}P statistical total correlation NMR spectroscopy of intact liver for metabolic biomarker assignment: application to galactosamine-induced hepatotoxicity, *Analytical Chemistry*, **2007**, 79(23): 8956-8966.
- Cole, D.E.C., Farag, S. and Dooley, K.C., Ethanolaminuria: a non-specific laboratory finding in the seriously ill infant, *Clinical Biochemistry*, **1988**, 21(5): 297-300.
- Constantinou, M.A., Papakonstantinou, E., Benaki, D., Spraul, M., Shulpis, K., Koupparis, M.A. and Mikros, E., Application of nuclear magnetic resonance spectroscopy combined with principal component analysis in detecting inborn errors of metabolism using blood spots: a metabonomic approach, *Analytica Chimica Acta*, **2004**, 511(2): 303-312.
- Constantinou, M.A., Papakonstantinou, E., Spraul, M., Sevastiadou, S., Costalos, C., Koupparis, M.A., Shulpis, K., Tsantili-Kakoulidou, A. and Mikros, E., ^1H NMR-based metabonomics for the diagnosis of inborn errors of metabolism in urine, *Analytica Chimica Acta*, **2005**, 542(2): 169-177.
- Craig, A., Cloarec, O., Holmes, E., Nicholson, J.K. and Lindon, J.C., Scaling and normalization effects in NMR spectroscopic metabonomic data sets, *Analytical Chemistry*, **2006**, 78(7): 2262-2267.
- Creasy, R.K., Resnik, R. and Iams, J.D., Creasy and Resnik's maternal-fetal medicine: principles and practice, **2009**, Philadelphia, Elsevier Saunders.
- Crockford, D.J., Holmes, E., Lindon, J.C., Plumb, R.S., Zirah, S., Bruce, S.J., Rainville, P., Stumpf, C.L. and Nicholson, J.K., Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: application in metabonomic toxicology studies, *Analytical Chemistry*, **2006**, 78(2): 363-371.
- Croze, M.L. and Soulage, C.O., Potential role and therapeutic interests of myo-inositol in metabolic diseases, *Biochimie*, **2013**, 95(10): 1811-1827.
- Cui, X.Q. and Churchill, G.A., Statistical tests for differential expression in cDNA microarray experiments, *Genome Biology*, **2003**, 4(4): 210.
- Culeddu, N., Chessa, M., Porcu, M., Fresu, P., Tonolo, G., Virgilio, G. and Migaleddu, V., NMR-based metabolomic study of type 1 diabetes, *Metabolomics*, **2012**, 8(6): 1162-1169.
- Dani, C., Bresci, C., Berti, E., Ottanelli, S., Mello, G., Mecacci, F., Breschi, R., Hu, X., Tenori, L. and Luchinat, C., Metabolomic profile of term infants of gestational diabetic mothers, *Journal of Maternal-Fetal and Neonatal Medicine*, **2013**, 1-6.
- Dasarathy, J., Gruca, L.L., Bennett, C., Parimi, P.S., Duenas, C., Marczewski, S., Fierro, J.L. and Kalhan, S.C., Methionine metabolism in human pregnancy, *The American Journal of Clinical Nutrition*, **2010**, 91(2): 357-365.
- DeFeo, E.M., Wu, C.-L., McDougal, W.S. and Cheng, L.L., A decade in prostate cancer: from NMR to metabolomics, *Nature Reviews Urology*, **2011**, 8(6): 301-311.
- Deja, S., Barg, E., Młynarz, P., Basiak, A. and Willak-Janc, E., ^1H NMR-based metabolomics studies of urine reveal differences between type 1 diabetic patients with high and low HbA_{1c} values, *Journal of Pharmaceutical and Biomedical Analysis*, **2013**, 83: 43-48.
- Dénes, J., Szabó, E., Robinette, S.L., Szatmári, I., Szőnyi, L., Kreuder, J.G., Rauterberg, E.W. and Takáts, Z., Metabonomics of newborn screening dried blood spot samples: A novel approach in the screening and diagnostics of inborn errors of metabolism, *Analytical Chemistry*, **2012**, 84(22): 10113-10120.
-

- Dessì, A., Atzori, L., Noto, A., Adriaan Visser, G.H., Gazzolo, D., Zanardo, V., Barberini, L., Puddu, M., Ottonello, G., Atzei, A., Magistris, A.D., Lussu, M., Murgia, F. and Fanos, V., Metabolomics in newborns with intrauterine growth retardation (IUGR): Urine reveals markers of metabolic syndrome, *Journal of Maternal-Fetal and Neonatal Medicine*, **2011**, 24(S2): 35-39.
- Dessì, A., Ottonello, G. and Fanos, V., Physiopathology of intrauterine growth retardation: From classic data to metabolomics, *Journal of Maternal-Fetal and Neonatal Medicine*, **2012**, 25(S5): 13-18.
- Di Anibal, C.V., Callao, M.P. and Ruisánchez, I., ¹H NMR variable selection approaches for classification. A case study: The determination of adulterated foodstuffs, *Talanta*, **2011**, 86: 316-323.
- Diaz, S.O., Pinto, J., Graça, G., Duarte, I.F., Barros, A.S., Galhano, E., Pita, C., Almeida, M.C., Goodfellow, B.J., Carreira, I.M. and Gil, A.M., Metabolic biomarkers of prenatal disorders: An exploratory NMR metabonomics study of second trimester maternal urine and blood plasma, *Journal of Proteome Research*, **2011**, 10(8): 3732-3742.
- Diaz, S.O., Barros, A.S., Goodfellow, B.J., Duarte, I.F., Carreira, I.M., Galhano, E., Pita, C., Almeida, M.d.C. and Gil, A.M., Following healthy pregnancy by nuclear magnetic resonance (NMR) metabolic profiling of human urine, *Journal of Proteome Research*, **2012**, 12(2): 969-979.
- Diaz, S.O., Barros, A.S., Goodfellow, B.J., Duarte, I.F., Galhano, E., Pita, C., Almeida, M.d.C., Carreira, I.M. and Gil, A.M., Second trimester maternal urine for the diagnosis of trisomy 21 and prediction of poor pregnancy outcomes, *Journal of Proteome Research*, **2013**, 12(6): 2946-2957.
- Dieterle, F., Ross, A., Schlotterbeck, G. and Senn, H., Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ¹H NMR metabonomics, *Analytical Chemistry*, **2006**, 78(13): 4281-4290.
- Dong, J., Cheng, K.-K., Xu, J., Chen, Z. and Griffin, J.L., Group aggregating normalization method for the preprocessing of NMR-based metabolomic data, *Chemometrics and Intelligent Laboratory Systems*, **2011**, 108(2): 123-132.
- Dou, L., Jourde-Chiche, N., Faure, V., Cerini, C., Berland, Y., Dignat-George, F. and Brunet, P., The uremic solute indoxyl sulfate induces oxidative stress in endothelial cells, *Journal of Thrombosis and Haemostasis*, **2007**, 5(6): 1302-1308.
- Duarte, I.F., Diaz, S.O. and Gil, A.M., NMR metabolomics of human blood and urine in disease research, *Journal of Pharmaceutical and Biomedical Analysis*, **2013**, <http://dx.doi.org/10.1016/j.jpba.2013.09.025>.
- Duarte, I.F. and Gil, A.M., Metabolic signatures of cancer unveiled by NMR spectroscopy of human biofluids, *Progress in Nuclear Magnetic Resonance Spectroscopy*, **2012**, 62: 51-74.
- Duarte, I.F., Legido-Quigley, C., Parker, D.A., Swann, J.R., Spraul, M., Braumann, U., Gil, A.M., Holmes, E., Nicholson, J.K., Murphy, G.M., Vilca-Melendez, H., Heaton, N. and Lindon, J.C., Identification of metabolites in human hepatic bile using 800 MHz ¹H NMR spectroscopy, HPLC-NMR/MS and UPLC-MS, *Molecular BioSystems*, **2009**, 5(2): 180-190.
- Dungan, K.M., Buse, J.B., Largay, J., Kelly, M.M., Button, E.A., Kato, S. and Wittlin, S., 1,5-Anhydroglucitol and postprandial hyperglycemia as measured by continuous

- glucose monitoring system in moderately controlled patients with diabetes, *Diabetes Care*, **2006**, 29(6): 1214-1219.
- Dunn, W., Brown, M., Worton, S., Davies, K., Jones, R., Kell, D. and Heazell, A., The metabolome of human placental tissue: investigation of first trimester tissue and changes related to preeclampsia in late pregnancy, *Metabolomics*, **2012**, 8(4): 579-597.
- Dunn, W.B., Broadhurst, D.I., Atherton, H.J., Goodacre, R. and Griffin, J.L., Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy, *Chemical Society Reviews*, **2011**, 40(1): 387-426.
- Dunn, W.B., Brown, M., Worton, S.A., Crocker, I.P., Broadhurst, D., Horgan, R., Kenny, L.C., Baker, P.N., Kell, D.B. and Heazell, A.E.P., Changes in the metabolic footprint of placental explant-conditioned culture medium identifies metabolic disturbances related to hypoxia and pre-eclampsia, *Placenta*, **2009**, 30(11): 974-980.
- Ellis, J., Athersuch, T., Thomas, L., Teichert, F., Perez-Trujillo, M., Svendsen, C., Spurgeon, D., Singh, R., Jarup, L., Bundy, J. and Keun, H., Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population, *BMC Medicine*, **2012**, 10(1): 61.
- Emwas, A.-H., Salek, R., Griffin, J. and Merzaban, J., NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations, *Metabolomics*, **2013**, 1-25.
- Enea, C., Seguin, F., Petitpas-Mulliez, J., Boildieu, N., Boisseau, N., Delpech, N., Diaz, V., Eugene, M. and Dugue, B., ¹H NMR-based metabolomics approach for exploring urinary metabolome modifications after acute and chronic physical exercise, *Analytical and Bioanalytical Chemistry*, **2010**, 396(3): 1167-1176.
- Fan, G., Wu, Z., Chen, L., Guo, Q., Ye, B. and Mao, J., Hypoxia-ischemic encephalopathy in full-term neonate: correlation proton MR spectroscopy with MR imaging, *European Journal of Radiology*, **2003**, 45(2): 91-98.
- Fan, T.W.M., Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures, *Progress in Nuclear Magnetic Resonance Spectroscopy*, **1996**, 28161-219.
- Fanos, V., Locci, E., Noto, A., Lazzarotto, T., Manzoni, P., Atzori, L. and Lanari, M., Urinary metabolomics in newborns infected by human cytomegalovirus: a preliminary investigation, *Early Human Development*, **2013**, 89 (Supplement 1): S58-61.
- Favé, G., Beckmann, M., Lloyd, A., Zhou, S., Harold, G., Lin, W., Taillart, K., Xie, L., Draper, J. and Mathers, J., Development and validation of a standardized protocol to monitor human dietary exposure by metabolite fingerprinting of urine samples, *Metabolomics*, **2011**, 7(4): 469-484.
- Favretto, D., Cosmi, E., Ragazzi, E., Visentin, S., Tucci, M., Fais, P., Cecchetto, G., Zanardo, V., Viel, G. and Ferrara, S.D., Cord blood metabolomic profiling in intrauterine growth restriction, *Analytical and Bioanalytical Chemistry*, **2012**, 402(3): 1109-1121.
- Fawcett, T., An introduction to ROC analysis, *Pattern Recognition Letters*, **2006**, 27(8): 861-874.
- Fiehn, O., Metabolomics - the link between genotypes and phenotypes, *Plant Molecular Biology*, **2002**, 48(1-2): 155-171.

-
- Foxall, P.J.D., Bewley, S., Neild, G.H., Rodeck, C.H. and Nicholson, J.K., Analysis of fetal and neonatal urine using proton nuclear magnetic resonance spectroscopy, *Archives of Disease in Childhood*, **1995**, 73(3): F153-F157.
- Freeman, D.J., Effects of maternal obesity on fetal growth and body composition: implications for programming and future health, *Seminars in Fetal and Neonatal Medicine*, **2010**, 15(2): 113-118.
- Friedrich, N., Metabolomics in diabetes research, *Journal of Endocrinology*, **2012**, 215(1): 29-42.
- Funkhouser, L.J. and Bordenstein, S.R., Mom knows best: The universality of maternal microbial transmission, *PLoS Biol*, **2013**, 11(8): e1001631.
- Gallery, E.D.M., Ross, M. and Gyory, A.Z., 24-Hour urinary creatinine excretion is not altered in human pregnancy, *Hypertension in Pregnancy*, **1996**, 15(2): 257-261.
- Gimenez, M., Soria-Pastor, S., Junque, C., Caldu, X., Narberhaus, A., Botet, F., Bargallo, N., Falcon, C. and Mercader, J.M., Proton magnetic resonance spectroscopy reveals medial temporal metabolic abnormalities in adolescents with history of preterm birth, *Pediatric Research*, **2008**, 64(5): 572-577.
- Glew, R.H., Melah, G., El-Nafaty, A.I., Brandt, Y., Morris, D. and VanderJagt, D.J., Plasma and urinary free amino acid concentrations in preeclamptic women in northern Nigeria, *Clinica Chimica Acta*, **2004**, 342(1-2): 179-185.
- Graça, G., *Metabonomics of human amniotic fluid for prenatal diagnostics*. PhD Thesis, **2013**, University of Aveiro.
- Graça, G., Diaz, S.O., Pinto, J., Barros, A.S., Duarte, I.F., Goodfellow, B.J., Galhano, E., Pita, C., Almeida, M.d.C., Carreira, I.M. and Gil, A.M., Can biofluids metabolic profiling help to improve healthcare during pregnancy?, *Spectroscopy: An International Journal*, **2012**, 27(5-6): 515-523.
- Graça, G., Duarte, I.F., Barros, A.S., Goodfellow, B.J., Diaz, S., Carreira, I.M., Couceiro, A.B., Galhano, E. and Gil, A.M., ¹H NMR based metabonomics of human amniotic fluid for the metabolic characterization of fetus malformations, *Journal of Proteome Research*, **2009**, 8(8): 4144-4150.
- Graça, G., Duarte, I.F., Barros, A.S., Goodfellow, B.J., Diaz, S.O., Pinto, J., Carreira, I.M., Galhano, E., Pita, C. and Gil, A.M., Impact of prenatal disorders on the metabolic profile of second trimester amniotic fluid: A nuclear magnetic resonance metabonomic study, *Journal of Proteome Research*, **2010**, 9(11): 6016-6024.
- Graça, G., Goodfellow, B.J., Barros, A.S., Diaz, S., Duarte, I.F., Spagou, K., Veselkov, K., Want, E.J., Lindon, J.C., Carreira, I.M., Galhano, E., Pita, C. and Gil, A.M., UPLC-MS metabolic profiling of second trimester amniotic fluid and maternal urine and comparison with NMR spectral profiling for the identification of pregnancy disorder biomarkers, *Molecular BioSystems*, **2012**, 8(4): 1243-1254.
- Graça, G., Moreira, A.S., Correia, A.J.V., Goodfellow, B.J., Barros, A.S., Duarte, I.F., Carreira, I.M., Galhano, E., Pita, C., Almeida, M.d.C. and Gil, A.M., Mid-infrared (MIR) metabolic fingerprinting of amniotic fluid: A possible avenue for early diagnosis of prenatal disorders?, *Analytica Chimica Acta*, **2013**, 764: 24-31.
- Gracie, S., Pennell, C., Ekman-Ordeberg, G., Lye, S., McManaman, J., Williams, S., Palmer, L., Kelley, M., Menon, R., Gravett, M. and Group, t.P.-O.R., An integrated systems biology approach to the study of preterm birth using "-omic" technology - a guideline for research, *BMC Pregnancy and Childbirth*, **2011**, 11(1): 71.
- Grenache, D.G. and Gronowski, A.M., Fetal lung maturity, *Clinical Biochemistry*, **2006**, 39(1): 1-10.
-

- Griffin, J.L., Atherton, H., Shockcor, J. and Atzori, L., Metabolomics as a tool for cardiac research, *Nature Reviews Cardiology*, **2011**, 8(11): 630-643.
- Groenen, P.M.W., Engelke, U.F., Wevers, R.A., Hendriks, J.C.M., Eskes, T.K.A.B., Merkus, H.M.W.M. and Steegers-Theunissen, R.P.M., High-resolution ¹H NMR spectroscopy of amniotic fluids from spina bifida fetuses and controls, *European Journal of Obstetrics & Gynecology and Reproductive Biology*, **2004**, 112(1): 16-23.
- Gucuyener, K., Atalay, Y., Aral, Y.Z., Hasanoglu, A., Turkyilmaz, C. and Biberoglu, G., Excitatory amino acids and taurine levels in cerebrospinal fluid of hypoxic ischemic encephalopathy in newborn, *Clinical Neurology and Neurosurgery*, **1999**, 101(3): 171-174.
- Hacker, N.F., Moore, J.G. and Gambone, J.C., Essentials of obstetrics and gynecology, **2004**, Philadelphia, Elsevier Saunders.
- Hadden, D.R. and McLaughlin, C., Normal and abnormal maternal metabolism during pregnancy, *Seminars in Fetal & Neonatal Medicine*, **2009**, 14(2): 66-71.
- Hashimoto, F., Nishiumi, S., Miyake, O., Takeichi, H., Chitose, M., Ohtsubo, H., Ishimori, S., Ninchoji, T., Hashimura, Y., Kaito, H., Morisada, N., Morioka, I., Fukuoka, H., Yoshida, M. and Iijima, K., Metabolomics analysis of umbilical cord blood clarifies changes in saccharides associated with delivery method, *Early Human Development*, **2013**, 89(5): 315-320.
- Hassan-Smith, G., Wallace, G.R., Douglas, M.R. and Sinclair, A.J., The role of metabolomics in neurological disease, *Journal of Neuroimmunology*, **2012**, 248(1-2): 48-52.
- Heazell, A.E.P., Bernatavicius, G., Warrander, L., Brown, M.C. and Dunn, W.B., A metabolomic approach identifies differences in maternal serum in third trimester pregnancies that end in poor perinatal outcome, *Reproductive Sciences*, **2012**, 19(8): 863-875.
- Heinzmann, S.S., Merrifield, C.A., Rezzi, S., Kochhar, S., Lindon, J.C., Holmes, E. and Nicholson, J.K., Stability and robustness of human metabolic phenotypes in response to sequential food challenges, *Journal of Proteome Research*, **2012**, 11(2): 643-655.
- Herrera, E., Lipid metabolism in pregnancy and its consequences in the fetus and newborn, *Endocrine*, **2002**, 19(1): 43-55.
- Herrera, E. and Ortega-Senovilla, H., Disturbances in lipid metabolism in diabetic pregnancy – Are these the cause of the problem?, *Best Practice & Research Clinical Endocrinology & Metabolism*, **2010**, 24(4): 515-525.
- Hod, M., Renzo, G.C.D., Leiva, A.D. and Langer, O., Textbook of diabetes and pregnancy, **2008**, London, Informa Healthcare.
- Hoffmann, E. and Stroobant, V., Mass Spectrometry: Principles and Applications, **2007**, London, John Wiley & Sons.
- Hogeveen, M., den Heijer, M., Semmekrot, B.A., Sporken, J.M., Ueland, P.M. and Blom, H.J., Umbilical choline and related methylamines betaine and dimethylglycine in relation to birth weight, *Pediatric Research*, **2013**, 73(6): 783-787.
- Holmes, E., Loo, R.L., Stamler, J., Bictash, M., Yap, I.K.S., Chan, Q., Ebbels, T., De Iorio, M., Brown, I.J., Veselkov, K.A., Daviglus, M.L., Kesteloot, H., Ueshima, H., Zhao, L., Nicholson, J.K. and Elliott, P., Human metabolic phenotype diversity and its association with diet and blood pressure, *Nature*, **2008**, 453(7193): 396-400.

-
- Horgan, R.P., Broadhurst, D.I., Dunn, W.B., Brown, M., Heazell, A.E.P., Kell, D.B., Baker, P.N. and Kenny, L.C., Changes in the metabolic footprint of placental explant-conditioned medium cultured in different oxygen tensions from placentas of small for gestational age and normal pregnancies, *Placenta*, **2010**, 31(10): 893-901.
- Horgan, R.P., Broadhurst, D.I., Walsh, S.K., Dunn, W.B., Brown, M., Roberts, C.T., North, R.A., McCowan, L.M., Kell, D.B., Baker, P.N. and Kenny, L.C., Metabolic profiling uncovers a phenotypic signature of small for gestational age in early pregnancy, *Journal of Proteome Research*, **2011**, 10(8): 3660-3673.
- Hourrier, S., Salomon, L.J., Dreux, S. and Muller, F., Screening for adverse pregnancy outcome at early gestational age, *Clinica Chimica Acta*, **2010**, 411(21-22): 1547-1552.
- Huang, S.T., Shu, K.H., Cheng, C.H., Wu, M.J., Yu, T.M., Chuang, Y.W. and Chen, C.H., Serum total *p*-cresol and indoxyl sulfate correlated with stage of chronic kidney disease in renal transplant recipients, *Transplantation Proceedings*, **2012**, 44(3): 621-624.
- Hyde, M.J., Griffin, J.L., Herrera, E., Byrne, C.D., Clarke, L. and Kemp, P.R., Delivery by Caesarean section, rather than vaginal delivery, promotes hepatic steatosis in piglets, *Clinical Science*, **2010**, 118(1-2): 47-59.
- Hyde, M.J., Mostyn, A., Modi, N. and Kemp, P.R., The health implications of birth by Caesarean section, *Biological Reviews*, **2012**, 87(1): 229-243.
- Hytten, F.E., The renal excretion of nutrients in pregnancy, *Postgraduate Medical Journal*, **1973**, 49(575): 625-629.
- Ivorra, C., Garcia-Vicent, C., Chaves, F.J., Monleon, D., Morales, J.M. and Lurbe, E., Metabolomic profiling in blood from umbilical cords of low birth weight newborns, *Journal of Translational Medicine*, **2012**, 10142.
- Jacobsen, N.E., NMR spectroscopy explained, **2007**, New Jersey, John Wiley & Sons.
- Jain, S., Jayasimhulu, K. and Clark, J.F., Metabolomic analysis of molecular species of phospholipids from normotensive and preeclamptic human placenta electrospray ionization mass spectrometry, *Frontiers In Bioscience*, **2004**, 93167-3175.
- Jankevics, A., Liepinsh, E., Vilskersts, R., Grinberga, S., Pugovics, O. and Dambrova, M., Metabolomic studies of experimental diabetic urine samples by ¹H NMR spectroscopy and LC/MS method, *Chemometrics and Intelligent Laboratory Systems*, **2008**, 97(1): 11-17.
- Kalhan, S.C. and Marczewski, S.E., Methionine, homocysteine, one carbon metabolism and fetal growth, *Reviews in Endocrine & Metabolic Disorders*, **2012**, 13(2): 109-119.
- Kaur, P., Rizk, N., Ibrahim, S., Luo, Y., Younes, N., Perry, B., Dennis, K., Zirrie, M., Luta, G. and Cheema, A.K., Quantitative metabolomic and lipidomic profiling reveals aberrant amino acid metabolism in type 2 diabetes, *Molecular BioSystems*, **2013**, 9(2): 307-317.
- Keeler, J., Understanding NMR Spectroscopy, **2002**, Cambridge, John Wiley & Sons, Ltd.
- Keller, P., Basic Principles of MR Imaging, **1988**, Milwaukee, GE, Medical Systems.
- Kenny, L.C., Broadhurst, D., Brown, M., Dunn, W.B., Redman, C.W.G., Kill, D.B. and Baker, P.N., Detection and identification of novel metabolomic biomarkers in preeclampsia, *Reproductive Sciences*, **2008**, 15(6): 591-597.
- Kenny, L.C., Broadhurst, D.I., Dunn, W., Brown, M., North, R.A., McCowan, L., Roberts, C., Cooper, G.J.S., Kell, D.B., Baker, P.N. and on behalf of the Screening for
-

- Pregnancy Endpoints Consortium, Robust early pregnancy prediction of later preeclampsia using metabolomic biomarkers, *Hypertension*, **2010**, 56(4): 741-749.
- Kenny, L.C., Dunn, W.B., Ellis, D.I., Myers, J., Bakera, P.N. and Kell, D.B., Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning, *Metabolomics*, **2005**, 1(3): 227-234.
- Keun, H.C., Sidhu, J., Pchejetski, D., Lewis, J.S., Marconell, H., Patterson, M., Bloom, S.R., Amber, V., Coombes, R.C. and Stebbing, J., Serum molecular signatures of weight change during early breast cancer chemotherapy, *Clinical Cancer Research*, **2009**, 15(21): 6716-6723.
- King, J.C., Physiology of pregnancy and nutrient metabolism, *The American Journal of Clinical Nutrition*, **2000**, 71(5): 1218S-1225S.
- Kliegman, R.M., Stanton, B.M.D., Geme, J.S., Schor, N. and Behrman, R.E., Nelson Textbook of Pediatrics, **2011**, Philadelphia, Elsevier Saunders.
- Kochhar, S., Jacobs, D.M., Ramadan, Z., Berruex, F., Fuerholz, A. and Fay, L.B., Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabolomics, *Analytical Biochemistry*, **2006**, 352(2): 274-281.
- Kohl, S.M., Klein, M.S., Hochrein, J., Oefner, P.J., Spang, R. and Gronwald, W., State-of-the-art data normalization methods improve NMR-based metabolomic analysis, *Metabolomics*, **2012**, 8(1): S146-S160.
- Kumps, A., Duez, P. and Mardens, Y., Metabolic, Nutritional, Iatrogenic, and Artfactual Sources of Urinary Organic Acids: A Comprehensive Table, *Clinical Chemistry*, **2002**, 48(5): 708-717.
- Lage, S., Andrade, F., Prieto, J.A., Asla, I., Rodriguez, A., Ruiz, N., Echeverria, J., Couce, M.L., Sanjurjo, P. and Aldamiz-Echevarria, L., Arginine-guanidinoacetate-creatine pathway in preterm newborns: creatine biosynthesis in newborns, *Journal of Pediatric Endocrinology & Metabolism*, **2013**, 26(1-2): 53-60.
- Lamarre, S.G., Morrow, G., Macmillan, L., Brosnan, M.E. and Brosnan, J.T., Formate: an essential metabolite, a biomarker, or more?, *Clinical Chemistry and Laboratory Medicine*, **2013**, 51(3): 571-578.
- Lanza, I.R., Zhang, S., Ward, L.E., Karakelides, H., Raftery, D. and Nair, K.S., Quantitative Metabolomics by ¹H-NMR and LC-MS/MS Confirms Altered Metabolic Pathways in Diabetes, *Plos One*, **2010**, 5(5): e10538.
- Lehtonen, H.-M., Lindstedt, A., Järvinen, R., Sinkkonen, J., Graça, G., Viitanen, M., Kallio, H. and Gil, A.M., ¹H NMR-based metabolic fingerprinting of urine metabolites after consumption of lingonberries (*Vaccinium vitis-idaea*) with a high-fat meal, *Food Chemistry*, **2013**, 138(2-3): 982-990.
- Lei, Z., Huhman, D.V. and Sumner, L.W., Mass Spectrometry Strategies in Metabolomics, *Journal of Biological Chemistry*, **2011**, 286(29): 25435-25442.
- Lenz, E.M., Bright, J., Wilson, I.D., Hughes, A., Morrisson, J., Lindberg, H. and Lockton, A., Metabolomics, dietary influences and cultural differences: a ¹H NMR-based study of urine samples obtained from healthy British and Swedish subjects, *Journal of Pharmaceutical and Biomedical Analysis*, **2004**, 36(4): 841-849.
- Lenz, E.M. and Wilson, I.D., Analytical strategies in metabolomics, *Journal of Proteome Research*, **2007**, 6: 443-458.
- Liang, X., Wang, Y., Liang, Q., Wang, Y., Huang, M. and Luo, G., Pathogenesis of neural tube defects: the story beyond methylation or one-carbon unit metabolism, *Metabolomics*, **2012**, 8(5): 919-929.
-

-
- Lin, X.H., Wang, Q.C., Yin, P.Y., Tang, L., Tan, Y.X., Li, H., Yan, K. and Xu, G.W., A method for handling metabonomics data from liquid chromatography/mass spectrometry: combinational use of support vector machine recursive feature elimination, genetic algorithm and random forest for feature selection, *Metabolomics*, **2011**, 7(4): 549-558.
- Lindon, J., Holmes, E. and Nicholson, J., Metabonomics Techniques and Applications to Pharmaceutical Research & Development, *Pharmaceutical Research*, **2006**, 23(6): 1075-1088.
- Lindon, J.C., Holmes, E., Bollard, M.E., Stanley, E.G. and Nicholson, J.K., Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis, *Biomarkers*, **2004**, 9(1): 1-31.
- Lindon, J.C. and Nicholson, J.K., Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery, *TrAC-Trends in Analytical Chemistry*, **2008**, 27(3): 194-204.
- Lindon, J.C. and Nicholson, J.K., Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics, *Annual Review of Analytical Chemistry*, **2008**, 1(1): 45-69.
- Lindon, J.C., Nicholson, J.K. and Holmes, E., The handbook of metabonomics and metabolomics, **2007**, Oxford, Elsevier.
- Lindon, J.C., Nicholson, J.K., Holmes, E. and Everett, J.R., Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids, *Concepts in Magnetic Resonance*, **2000**, 12(5): 289-320.
- Lindon, J.C., Tranter, G.E. and Holmes, J.L., Encyclopedia of spectroscopy and spectrometry, **2000**, London, Academic Press.
- Llorach, R., Garcia-Aloy, M., Tulipani, S., Vazquez-Fresno, R. and Andres-Lacueva, C., Nutrimetabolomic strategies to develop new biomarkers of intake and health effects, *Journal of Agricultural and Food Chemistry*, **2012**, 60(36): 8797-8808.
- Lloyd, A.J., Favé, G., Beckmann, M., Lin, W., Taillart, K., Xie, L., Mathers, J.C. and Draper, J., Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods, *The American Journal of Clinical Nutrition*, **2011**, 94(4): 981-991.
- Lutz, N.W. and Cozzone, P.J., Metabolic profiling in multiple sclerosis and other disorders by quantitative analysis of cerebrospinal fluid using nuclear magnetic resonance spectroscopy, *Current Pharmaceutical Biotechnology*, **2011**, 12(7): 1016-1025.
- Ma, S., Shieh, L.-I. and Huang, C.-C., High-resolution proton nuclear magnetic resonance studies of urine from asphyxiated newborn infants, *Applied Biochemistry and Biotechnology*, **1995**, 53(1): 37-51.
- Madsen, R., Lundstedt, T. and Trygg, J., Chemometrics in metabolomics- A review in human disease diagnosis, *Analytica Chimica Acta*, **2009**, 659(2010): 23-33.
- Maher, A.D., Cysique, L.A., Brew, B.J. and Rae, C.D., Statistical integration of ^1H NMR and MRS data from different biofluids and tissues enhances recovery of biological information from individuals with HIV-1 infection, *Journal of Proteome Research*, **2011**, 10(4): 1737-1745.
- Makinen, V.P., Forsblom, C., Thorn, L.M., Waden, J., Gordin, D., Heikkila, O., Hietala, K., Kyllonen, L., Kyto, J., Rosengard-Barlund, M., Saraheimo, M., Tolonen, N., Parkkonen, M., Kaski, K., Ala-Korpela, M. and Groop, P.H., Metabolic phenotypes, vascular complications, and premature deaths in a population of 4,197 patients with type 1 diabetes, *Diabetes*, **2008**, 57(9): 2480-2487.
-

- Makinen, V.P., Soininen, P., Forsblom, C., Parkkonen, M., Ingman, P., Kaski, K., Groop, P.H., Ala-Korpela, M. and FinnDiane Study, G., ^1H NMR metabonomics approach to the disease continuum of diabetic complications and premature death, *Molecular Systems Biology*, **2008**, 4(1): 167.
- Malik, G.K., Pandey, M., Kumar, R., Chawla, S., Rath, B. and Gupta, R.K., MR imaging and in vivo proton spectroscopy of the brain in neonates with hypoxic ischemic encephalopathy, *European Journal of Radiology*, **2002**, 43(1): 6-13.
- Marescau, B., Nagels, G., Possemiers, I., De Broe, M.E., Becaes, I., Billiouw, J.-M., Lornoy, W. and De Deyn, P.P., Guanidino compounds in serum and urine of nondialyzed patients with chronic renal insufficiency, *Metabolism*, **1997**, 46(9): 1024-1031.
- Martin, F.P.J., Rezzi, S., Pere-Trepat, E., Kamlage, B., Collino, S., Leibold, E., Kastler, J., Rein, D., Fay, L.B. and Kochhar, S., Metabolic effects of dark chocolate consumption on energy, gut microbiota, and stress-related metabolism in free-living subjects, *Journal of Proteome Research*, **2009**, 8(12): 5568-5579.
- Meiboom, S. and Gill, D., Modified spin-echo method for measuring nuclear relaxation times, *Review of Scientific Instruments*, **1958**, 29(8): 688-691.
- Messana, I., Forni, F., Ferrari, F., Rossi, C., Giardina, B. and Zuppi, C., Proton nuclear magnetic resonance spectral profiles of urine in type II diabetic patients, *Clinical Chemistry*, **1998**, 44(7): 1529-1534.
- Micle, O., Muresan, M., Antal, L., Bodog, F. and Bodog, A., The influence of homocysteine and oxidative stress on pregnancy outcome, *Journal of Medicine & Life*, **2012**, 5(1): 68-73.
- Miller, S., Ruttinger, V. and Macy, I.G., Urinary excretion of ten amino acids by women during the reproductive cycle, *Journal of Biological Chemistry*, **1954**, 209(2): 795-801.
- Moore, K.L., Persaud, T.V.N. and Torchia, M.G., The developing human: clinically oriented embryology, **2008**, Philadelphia, Elsevier Saunders.
- Morelli, L., Postnatal Development of Intestinal Microflora as Influenced by Infant Nutrition, *The Journal of Nutrition*, **2008**, 138(9): 1791S-1795S.
- Morris, C.A., Introduction: Williams syndrome, *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, **2010**, 154C(2): 203-208.
- Nakagawa, S. and Cuthill, I.C., Effect size, confidence interval and statistical significance: a practical guide for biologists, *Biological Reviews*, **2007**, 82(4): 591-605.
- Nicholson, J.K., Everett, J.R. and Lindon, J.C., Longitudinal pharmacometabonomics for predicting patient responses to therapy: drug metabolism, toxicity and efficacy, *Expert Opinion on Drug Metabolism Toxicology*, **2012**, 8(2): 135-139.
- Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W. and Pettersson, S., Host-gut microbiota metabolic interactions, *Science*, **2012**, 336(6086): 1262-1267.
- Nicholson, J.K., Holmes, E., Kinross, J.M., Darzi, A.W., Takats, Z. and Lindon, J.C., Metabolic phenotyping in clinical and surgical environments, *Nature*, **2012**, 491(7424): 384-392.
- Nicholson, J.K., Lindon, J.C. and Holmes, E., 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data, *Xenobiotica*, **1999**, 29(11): 1181-1189.
-

-
- Nicholson, J.K., Wilson, I.D. and Lindon, J.C., Pharmacometabonomics as an effector for personalized medicine, *Pharmacogenomics*, **2011**, 12(1): 103-111.
- Nicolaidis, K.H., Screening for fetal aneuploidies at 11 to 13 weeks, *Prenatal Diagnosis*, **2011**, 31(1): 7-15.
- Nissen, P.M., Nebel, C., Oksbjerg, N. and Bertram, H.C., Metabolomics reveals relationship between plasma inositols and birth weight: Possible markers for fetal programming of type 2 diabetes, *Journal of Biomedicine and Biotechnology*, **2011**, 2011ID 378268.
- Obeid, R., Hartmuth, K., Herrmann, W., Gortner, L., Rohrer, T.R., Geisel, J., Reed, M.C. and Nijhout, H.F., Blood biomarkers of methylation in Down syndrome and metabolic simulations using a mathematical model, *Molecular Nutrition & Food Research*, **2012**, 56(10): 1582-1589.
- Odibo, A.O., Goetzinger, K.R., Odibo, L., Cahill, A.G., Macones, G.A., Nelson, D.M. and Dietzen, D.J., First-trimester prediction of preeclampsia using metabolomic biomarkers: a discovery phase study, *Prenatal Diagnosis*, **2011**, 31(10): 990-994.
- Ombro, D., Salvatore, F. and Ruoppolo, M., Quantitative liquid chromatography coupled with tandem mass spectrometry analysis of urinary acylglycines: Application to the diagnosis of inborn errors of metabolism, *Analytical Biochemistry*, **2011**, 417(1): 122-128.
- Oresic, M., Simell, S., Sysi-Aho, M., Nanto-Salonen, K., Seppanen-Laakso, T., Parikka, V., Katajamaa, M., Hekkala, A., Mattila, I., Keskinen, P., Yetukuri, L., Reinikainen, A., Lahde, J., Suortti, T., Hakalax, J., Simell, T., Hyoty, H., Veijola, R., Ilonen, J., Lahesmaa, R., Knip, M. and Simell, O., Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes, *Journal of Experimental Medicine*, **2008**, 205(13): 2975-2984.
- Ottolenghi, C., Abermil, N., Lescoat, A., Aupetit, J., Beaugendre, O., Morichon-Delvallez, N., Ricquier, D., Chadefaux-Vekemans, B. and Rabier, D., Gestational age-related reference values for amniotic fluid organic acids, *Prenatal Diagnosis*, **2010**, 30(1): 43-48.
- Pappa, K.I., Vlachos, G., Theodora, M., Roubelaki, M., Angelidou, K. and Antsaklis, A., Intermediate metabolism in association with the amino acid profile during the third trimester of normal pregnancy and diet-controlled gestational diabetes, *American Journal of Obstetrics and Gynecology*, **2007**, 196(1): 65.e61-65.e65.
- Perlman, M., Diagnosis and managing neonatal respiratory depression, *Canadian Family Physician*, **1985**, 31:1019-1023.
- Perrone, S., Mussap, M., Longini, M., Fanos, V., Bellieni, C.V., Proietti, F., Cataldi, L. and Buonocore, G., Oxidative kidney damage in preterm newborns during perinatal period, *Clinical Biochemistry*, **2007**, 40(9-10): 656-660.
- Petrie, A. and Sabin, C., Medical Statistics at a Glance, **2009**, Oxford, Blackwell Science, Ltd.
- Psihogios, N.G., Gazi, I.F., Elisaf, M.S., Seferiadis, K.I. and Bairaktari, E.T., Gender-related and age-related urinalysis of healthy subjects by NMR-based metabonomics, *NMR in Biomedicine*, **2008**, 21(3): 195-207.
- Qiu, Y., Rajagopalan, D., Connor, S.C., Damian, D., Zhu, L., Handzel, A., Hu, G., Amanullah, A., Bao, S., Woody, N., MacLean, D., Lee, K., Vanderwall, D. and Ryan, T., Multivariate classification analysis of metabolomic data for candidate biomarker discovery in type 2 diabetes mellitus, *Metabolomics*, **2008**, 4(4): 337-346.
-

- Quintás, G., Portillo, N., García-Cañaveras, J., Castell, J., Ferrer, A. and Lahoz, A., Chemometric approaches to improve PLS-DA model outcome for predicting human non-alcoholic fatty liver disease using UPLC-MS as a metabolic profiling tool, *Metabolomics*, **2012**, 8(1): 86-98.
- Ramautar, R., Nevedomskaya, E., Mayboroda, O.A., Deelder, A.M., Wilson, I.D., Gika, H.G., Theodoridis, G.A., Somsen, G.W. and de Jong, G.J., Metabolic profiling of human urine by CE-MS using a positively charged capillary coating and comparison with UPLC-MS, *Molecular BioSystems*, **2011**, 7(1): 194-199.
- Ratnam, S., Wijekoon, E.P., Hall, B., Garrow, T.A., Brosnan, M.E. and Brosnan, J.T., Effects of diabetes and insulin on betaine-homocysteine S-methyltransferase expression in rat liver, *American Journal of Physiology - Endocrinology And Metabolism*, **2006**, 290(5): E933-E939.
- Regnault, T.R.H., Friedman, J.E., Wilkening, R.B., Anthony, R.V. and Hay Jr, W.W., Fetoplacental transport and utilization of amino acids in IUGR — a review, *Placenta*, **2005**, 26(Supplement): S52-S62.
- Reinke, S.N., Walsh, B.H., Boylan, G.B., Sykes, B.D., Kenny, L.C., Murray, D.M. and Broadhurst, D.I., ¹H NMR derived metabolomic profile of neonatal asphyxia in umbilical cord serum: Implications for hypoxic ischemic encephalopathy, *Journal of Proteome Research*, **2013**, 12(9): 4230-4239.
- Rezzi, S., Ramadan, Z., Fay, L.B. and Kochhar, S., Nutritional metabonomics: Applications and perspectives, *Journal of Proteome Research*, **2007**, 6(2): 513-525.
- Rezzi, S., Ramadan, Z., Martin, F.P.J., Fay, L.B., van Bladeren, P., Lindon, J.C., Nicholson, J.K. and Kochhar, S., Human metabolic phenotypes link directly to specific dietary preferences in healthy individuals, *Journal of Proteome Research*, **2007**, 6(11): 4469-4477.
- Rivière, C., Thi Hong, V.N., Hoai, N.N., Dejaegher, B., Tistaert, C., Van, K.P., Heyden, Y.V., Chau Van, M. and Quetin-Leclercq, J., N-methyl-5-carboxamide-2-pyridone from *Mallotus barbatus*: A chemosystematic marker of the Euphorbiaceae genus *Mallotus*, *Biochemical Systematics and Ecology*, **2012**, 44: 212-215.
- Rocha, C.M., Barros, A.S., Gil, A.M., Goodfellow, B.J., Humpfer, E., Spraul, M., Carreira, I.M., Melo, J.B., Bernardo, J., Gomes, A., Sousa, V., Carvalho, L. and Duarte, I.F., Metabolic profiling of human lung cancer tissue by ¹H high resolution magic angle spinning (HRMAS) NMR spectroscopy, *Journal of Proteome Research*, **2010**, 9(1): 319-332.
- Roldan, A., Figueras-Aloy, J., Deulofeu, R. and Jimenez, R., Glycine and other neurotransmitter amino acids in cerebrospinal fluid in perinatal asphyxia and neonatal hypoxic-ischaemic encephalopathy, *Acta Paediatrica*, **1999**, 88(10): 1137-1141.
- Romero, R., Mazaki-Tovi, S., Vaisbuch, E., Kusanovic, J.P., Chaiworapongsa, T., Gomez, R., Nien, J.K., Yoon, B.H., Mazar, M., Luo, J.Q., Banks, D., Ryals, J. and Beecher, C., Metabolomics in premature labor: a novel approach to identify patients at risk for preterm delivery, *Journal of Maternal-Fetal & Neonatal Medicine*, **2010**, 23(12): 1344-1359.
- Roux, A., Lison, D., Junot, C. and Heilier, J.F., Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review, *Clinical Biochemistry*, **2011**, 44(1): 119-135.

-
- Saccenti, E., Hoefsloot, H.J., Smilde, A., Westerhuis, J. and Hendriks, M.W.B., Reflections on univariate and multivariate analysis of metabolomics data, *Metabolomics*, **2013**, 1-14.
- Sachse, D., Sletner, L., Mørkrid, K., Jenum, A.K., Birkeland, K.I., Rise, F., Piehler, A.P. and Berg, J.P., Metabolic changes in urine during and after pregnancy in a large, multiethnic population-based cohort study of gestational diabetes, *Plos One*, **2012**, 7(12): e52399.
- Salek, R.M., Maguire, M.L., Bentley, E., Rubtsov, D.V., Hough, T., Cheeseman, M., Nunez, D., Sweatman, B.C., Haselden, J.N., Cox, R.D., Connor, S.C. and Griffin, J.L., A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human, *Physiological Genomics*, **2007**, 29(2): 99-108.
- Saude, E.J., Adamko, D., Rowe, B.H., Marrie, T. and Sykes, B.D., Variation of metabolites in normal human urine, *Metabolomics*, **2007**, 3439-451.
- Saude, E.J., Skappak, C.D., Regush, S., Cook, K., Ben-Zvi, A., Becker, A., Moqbel, R., Sykes, B.D., Rowe, B.H. and Adamko, D.J., Metabolomic profiling of asthma: Diagnostic utility of urine nuclear magnetic resonance spectroscopy, *Journal of Allergy and Clinical Immunology*, **2011**, 127(3): 757-764.e756.
- Savorani, F., Tomasi, G. and Engelsen, S.B., icoshift: A versatile tool for the rapid alignment of 1D NMR spectra, *Journal of Magnetic Resonance*, **2010**, 202(2): 190-202.
- Schicho, R., Shaykhutdinov, R., Ngo, J., Nazyrova, A., Schneider, C., Panaccione, R., Kaplan, G.G., Vogel, H.J. and Storr, M., Quantitative metabolomic profiling of serum, plasma, and urine by ¹H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals, *Journal of Proteome Research*, **2012**, 11(6): 3344-3357.
- Scholtens, D.M., Muehlbauer, M.J., Daya, N.R., Stevens, R.D., Dyer, A.R., Lowe, L.P., Metzger, B.E., Newgard, C.B., Bain, J.R., Lowe, W.L. and for the HAPO Study Cooperative Research Group, Metabolomics reveals broad-scale metabolic perturbations in hyperglycemic mothers during pregnancy, *Diabetes Care*, **2013**, doi:10.2337/dc13-0989.
- Scott, J.R., Gibbs, R.S., Karlan, B.Y., Haney, A.F. and Danforth, D.N., *Danforth's Obstetrics and Gynecology*, **2003**, Philadelphia, Lippincott Williams & Wilkins Publishers.
- Senn, T., Hazen, S.L. and Tang, W.H.W., Translating Metabolomics to Cardiovascular Biomarkers, *Progress in Cardiovascular Diseases*, **2012**, 55(1): 70-76.
- Shprintzen, R.J., Velo-cardio-facial syndrome: 30 years of study, *Developmental Disabilities Research Reviews*, **2008**, 14(1): 3-10.
- Slupsky, C.M., Rankin, K.N., Fu, H., Chang, D., Rowe, B.H., Charles, P.G.P., McGeer, A., Low, D., Long, R., Kunitomo, D., Sawyer, M.B., Fedorak, R.N., Adamko, D.J., Saude, E.J., Shah, S.L. and Marrie, T.J., Pneumococcal pneumonia: Potential for diagnosis through a urinary metabolic profile, *Journal of Proteome Research*, **2009**, 8(12): 5550-5558.
- Slupsky, C.M., Rankin, K.N., Wagner, J., Fu, H., Chang, D., Weljie, A.M., Saude, E.J., Lix, B., Adamko, D.J., Shah, S., Greiner, R., Sykes, B.D. and Marrie, T.J., Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles, *Analytical Chemistry*, **2007**, 79(18): 6995-7004.
-

- Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. and Siuzdak, G., XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification, *Analytical Chemistry*, **2006**, 78(3): 779-787.
- Smolinska, A., Blanchet, L., Buydens, L.M.C. and Wijmenga, S.S., NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review, *Analytica Chimica Acta*, **2012**, 750: 82-97.
- Solanky, K.S., Bailey, N.J., Beckwith-Hall, B.M., Bingham, S., Davis, A., Holmes, E., Nicholson, J.K. and Cassidy, A., Biofluid H-1 NMR-based metabonomic techniques in nutrition research metabolic effects of dietary isoflavones in humans, *Journal of Nutritional Biochemistry*, **2005**, 16(4): 236-244.
- Spagou, K., Wilson, I.D., Masson, P., Theodoridis, G., Raikos, N., Coen, M., Holmes, E., Lindon, J.C., Plumb, R.S., Nicholson, J.K. and Want, E.J., HILIC-UPLC-MS for exploratory urinary metabolic profiling in toxicological studies, *Analytical Chemistry*, **2011**, 83(1): 382-390.
- Stella, C., Beckwith-Hall, B., Cloarec, O., Holmes, E., Lindon, J.C., Powell, J., van der Ouderaa, F., Bingham, S., Cross, A.J. and Nicholson, J.K., Susceptibility of human metabolic phenotypes to dietary modulation, *Journal of Proteome Research*, **2006**, 5(10): 2780-2788.
- Struthers, J.L., Carson, N., McGill, M. and Khalifa, M.M., Molecular screening for Smith-Magenis syndrome among patients with mental retardation of unknown cause, *Journal of Medical Genetics*, **2002**, 39(10): e59.
- Stuart, E.L., Evans, G.S., Lin, Y.S. and Powers, H.J., Reduced collagen and ascorbic acid concentrations and increased proteolytic susceptibility with prelabor fetal membrane rupture in women, *Biology of Reproduction*, **2005**, 72(1): 230-235.
- Su, M., Zheng, X., Zhang, T., Pei, L., Wang, F., Zheng, X., Gu, X., Song, X., Lu, X., Chen, G., Bao, Y., Chen, T., Zhao, A., Bao, Y., Jia, W., Zeisel, S. and Jia, W., Integrated profiling of metabolites and trace elements reveals a multifaceted malnutrition in pregnant women from a region with a high prevalence of congenital malformations, *Metabolomics*, **2012**, 8(5): 831-844.
- Su, M.M., Zheng, X.Y., Zhang, T., Pei, L.J., Wang, F., Zheng, X.J., Gu, X., Song, X.M., Lu, X.L., Chen, G., Bao, Y.H., Chen, T.L., Zhao, A.H., Bao, Y.Q., Jia, W.P., Zeisel, S.H. and Jia, W., Integrated profiling of metabolites and trace elements reveals a multifaceted malnutrition in pregnant women from a region with a high prevalence of congenital malformations, *Metabolomics*, **2012**, 8(5): 831-844.
- Suhre, K., Meisinger, C., Doring, A., Altmaier, E., Belcredi, P., Gieger, C., Chang, D., Milburn, M.V., Gall, W.E., Weinberger, K.M., Mewes, H.W., de Angelis, M.H., Wichmann, H.E., Kronenberg, F., Adamski, J. and Illig, T., Metabolic footprint of diabetes: A multiplatform metabolomics study in an epidemiological setting, *Plos One*, **2010**, 5(11): e13953.
- Sun, X.-M., Yu, X.-P., Liu, Y., Xu, L. and Di, D.-L., Combining bootstrap and uninformative variable elimination: Chemometric identification of metabonomic biomarkers by nonparametric analysis of discriminant partial least squares, *Chemometrics and Intelligent Laboratory Systems*, **2012**, 115:37-43.
- Sussulini, A., Prando, A., Maretto, D.A., Poppi, R.J., Tasic, L., Muller Banzato, C.E. and Zezzi Arruda, M.A., Metabolic profiling of human blood serum from treated patients with bipolar disorder employing ¹H NMR spectroscopy and chemometrics, *Analytical Chemistry*, **2009**, 81(23): 9755-9763.
-

-
- Tapp, H.S. and Kemsley, E.K., Notes on the practical utility of OPLS, *TrAC Trends in Analytical Chemistry*, **2009**, 28(11): 1322-1327.
- Tax, N., Urlesberger, B., Binder, C., Pocivalnik, M., Morris, N. and Pichler, G., The influence of perinatal asphyxia on peripheral oxygenation and perfusion in neonates, *Early Human Development*, **2013**, 89(7): 483-486.
- Tea, I., Le Gall, G., Kuster, A., Guignard, N., Alexandre-Gouabau, M.-C., Darmaun, D. and Robins, R.J., ¹H NMR-based metabolic profiling of maternal and umbilical cord blood indicates altered materno-foetal nutrient exchange in preterm infants, *Plos One*, **2012**, 7(1): e29947.
- Tharapel, A.T., Kadandale, J.S., Martens, P.R., Wachtel, S.S. and Wilroy, R.S., Prader Willi/Angelman and DiGeorge/velocardiofacial syndrome deletions: Diagnosis by primed in situ labeling (PRINS), *American Journal of Medical Genetics*, **2002**, 107(2): 119-122.
- Theodoridis, G.A., Gika, H.G., Want, E.J. and Wilson, I.D., Liquid chromatography-mass spectrometry based global metabolite profiling: A review, *Analytica Chimica Acta*, **2012**, 7117-16.
- Thomas, E.L., Parkinson, J.R., Hyde, M.J., Yap, I.K.S., Holmes, E., Dore, C.J., Bell, J.D. and Modi, N., Aberrant adiposity and ectopic lipid deposition characterize the adult phenotype of the preterm infant, *Pediatric Research*, **2011**, 70(5): 507-512.
- Torgrip, R., Åberg, K., Alm, E., Schuppe-Koistinen, I. and Lindberg, J., A note on normalization of biofluid 1D ¹H-NMR data, *Metabolomics*, **2008**, 4(2): 114-121.
- Trump, S., Laudi, S., Unruh, N., Goelz, R. and Leibfritz, D., ¹H-NMR metabolic profiling of human neonatal urine, *Magnetic Resonance Materials in Physics Biology and Medicine*, **2006**, 19(6): 305-312.
- Trygg, J., Holmes, E. and Lundstedt, T., Chemometrics in Metabonomics, *Journal of Proteome Research*, **2007**, 6(2): 469-479.
- Turgan, N., Boydak, B., Habif, S., Gultur, C., Senol, B., Mutaf, I., Ozmen, D. and Bayindir, O., Urinary hypoxanthine and xanthine levels in acute coronary syndromes, *International Journal of Clinical & Laboratory Research*, **1999**, 29(4): 162-165.
- Turner, E., Brewster, J.A., Simpson, N.A.B., Walker, J.J. and Fisher, J., Plasma from women with preeclampsia has a low lipid and ketone body content—A nuclear magnetic resonance study, *Hypertension in Pregnancy*, **2007**, 26(3): 329-342.
- Turner, E., Brewster, J.A., Simpson, N.A.B., Walker, J.J. and Fisher, J., Aromatic amino acid biomarkers of preeclampsia - A nuclear magnetic resonance investigation, *Hypertension in Pregnancy*, **2008**, 27(3): 225-235.
- Turner, E., Brewster, J.A., Simpson, N.A.B., Walker, J.J. and Fisher, J., Imidazole-based erythrocyte markers of oxidative stress in preeclampsia--An NMR investigation, *Reproductive Sciences*, **2009**, 16(11): 1040-1051.
- Vakilian, K., Ranjbar, A., Zarganjfard, A., Mortazavi, M., Vosough-Ghanbari, S., Mashaiee, S. and Abdollahi, M., On the relation of oxidative stress in delivery mode in pregnant women; A toxicological concern, *Toxicology Mechanisms and Methods*, **2009**, 19(2): 94-99.
- van den Berg, R.A., Hoefsloot, H.C.J., Westerhuis, J.A., Smilde, A.K. and van der Werf, M.J., Centering, scaling, and transformations: improving the biological information content of metabolomics data, *BMC Genomics*, **2006**, 7(1): 142-142.
- van Doorn, M., Vogels, J., Tas, A., van Hoogdalem, E.J., Burggraaf, J., Cohen, A. and van der Greef, J., Evaluation of metabolite profiles as biomarkers for the
-

- pharmacological effects of thiazolidinediones in Type 2 diabetes mellitus patients and healthy volunteers, *British Journal of Clinical Pharmacology*, **2007**, 63(5): 562-574.
- van Kuilenburg, A.B.P., Stroomer, A.E.M., van Lenthe, H., Abeling, N. and van Gennip, A.H., New insights in dihydropyrimidine dehydrogenase deficiency: a pivotal role for beta-aminoisobutyric acid?, *Biochemical Journal*, **2004**, 379:119-124.
- van Velzen, E.J.J., Westerhuis, J.A., van Duynhoven, J.P.M., van Dorsten, F.A., Grun, C.H., Jacobs, D.M., Duchateau, G., Vis, D.J. and Smilde, A.K., Phenotyping tea consumers by nutrikinetic analysis of polyphenolic end-metabolites, *Journal of Proteome Research*, **2009**, 8(7): 3317-3330.
- Veselkov, K.A., Lindon, J.C., Ebbels, T.M.D., Crockford, D., Volynkin, V.V., Holmes, E., Davies, D.B. and Nicholson, J.K., Recursive segment-wise peak alignment of biological ^1H NMR spectra for improved metabolic biomarker recovery, *Analytical Chemistry*, **2009**, 81(1): 56-66.
- Veselkov, K.A., Vingara, L.K., Masson, P., Robinette, S.L., Want, E., Li, J.V., Barton, R.H., Boursier-Neyret, C., Walther, B., Ebbels, T.M., Pelczer, I.n., Holmes, E., Lindon, J.C. and Nicholson, J.K., Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery, *Analytical Chemistry*, **2011**, 83(15): 5864-5872.
- Villas-Bôas, S.G., Mas, S., Åkesson, M., Smedsgaard, J. and Nielsen, J., Mass spectrometry in metabolome analysis, *Mass Spectrometry Reviews*, **2005**, 24(5): 613-646.
- Vinaixa, M., Samino, S., Saez, I., Duran, J., Guinovart, J.J. and Yanes, O., A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data, *Metabolites*, **2012**, 2(4): 775-795.
- Vu, T.N., Valkenburg, D., Smets, K., Verwaest, K.A., Dommissie, R., Lemiere, F., Verschoren, A., Goethals, B. and Laukens, K., An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data, *BMC Bioinformatics*, **2011**, 12:405.
- Walsh, B.H., Broadhurst, D.I., Mandal, R., Wishart, D.S., Boylan, G.B., Kenny, L.C. and Murray, D.M., The metabolomic profile of umbilical cord blood in neonatal hypoxic ischaemic encephalopathy, *Plos One*, **2012**, 7(12): e50520.
- Wang, T.J., Larson, M.G., Vasan, R.S., Cheng, S., Rhee, E.P., McCabe, E., Lewis, G.D., Fox, C.S., Jacques, P.F., Fernandez, C., O'Donnell, C.J., Carr, S.A., Mootha, V.K., Florez, J.C., Souza, A., Melander, O., Clish, C.B. and Gerszten, R.E., Metabolite profiles and the risk of developing diabetes, *Nature Medicine*, **2011**, 17(4): 448-453.
- Wang, Y., Cloarec, O., Tang, H.R., Lindon, J.C., Holmes, E., Kochhar, S. and Nicholson, J.K., Magic angle spinning NMR and H-1-P-31 heteronuclear statistical total correlation spectroscopy of intact human gut biopsies, *Analytical Chemistry*, **2008**, 80(4): 1058-1066.
- Wang, Y.L., Tang, H.R., Nicholson, J.K., Hylands, P.J., Sampson, J. and Holmes, E., A metabonomic strategy for the detection of the metabolic effects of chamomile (*Matricaria recutita* L.) ingestion, *Journal of Agricultural and Food Chemistry*, **2005**, 53(2): 191-196.
- Want, E.J., Nordström, A., Morita, H. and Siuzdak, G., From exogenous to endogenous: The inevitable imprint of mass spectrometry in metabolomics, *Journal of Proteome Research*, **2006**, 6(2): 459-468.

-
- Want, E.J., Wilson, I.D., Gika, H., Theodoridis, G., Plumb, R.S., Shockcor, J., Holmes, E. and Nicholson, J.K., Global metabolic profiling procedures for urine using UPLC-MS, *Nature Protocols*, **2010**, 5(6): 1005-1018.
- Watanabe, K., Iwasaki, A., Mori, T., Kimura, C., Matsushita, H., Shinohara, K. and Wakatsuki, A., Differences in levels of oxidative stress in mothers and neonate: the impact of mode of delivery, *Journal of Maternal-Fetal and Neonatal Medicine*, **2013**, 26(16): 1649-1652.
- Weksberg, R., Shuman, C. and Beckwith, J.B., Beckwith-Wiedemann syndrome, *European Journal of Human Genetics*, **2010**, 18(1): 8-14.
- Westerhuis, J., Hoefsloot, H., Smit, S., Vis, D., Smilde, A., van Velzen, E., van Duijnhoven, J. and van Dorsten, F., Assessment of PLSDA cross validation, *Metabolomics*, **2008**, 4(1): 81-89.
- Wevers, R.A., Engelke, U.F.H., Moolenaar, S.H., Bräutigam, C., de Jong, J.G.N., Duran, R., de Abreu, R.A. and van Gennip, A.H., ¹H-NMR spectroscopy of body fluids: Inborn errors of purine and pyrimidine metabolism, *Clinical Chemistry*, **1999**, 45(4): 539-548.
- Wiklund, S., Nilsson, D., Eriksson, L., Sjöström, M., Wold, S. and Faber, K., A randomization test for PLS component selection, *Journal of Chemometrics*, **2007**, 21(10-11): 427-439.
- Wilcox, G., Insulin and insulin resistance, *The Clinical biochemist. Reviews / Australian Association of Clinical Biochemists*, **2005**, 26(2): 19-39.
- Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J.A., Lim, E., Sobsey, C.A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H.J. and Forsythe, I., HMDB: a knowledgebase for the human metabolome, *Nucleic Acids Research*, **2009**, 37(suppl 1): D603-D610.
- Wold, S., Sjöström, M. and Eriksson, L., PLS-regression: a basic tool of chemometrics, *Chemometrics and Intelligent Laboratory Systems*, **2001**, 58(2): 109-130.
- Woods Jr, J.R., Plessinger, M.A. and Miller, R.K., Vitamins C and E: Missing links in preventing preterm premature rupture of membranes?, *American Journal of Obstetrics and Gynecology*, **2001**, 185(1): 5-10.
- Wurtz, P., Soininen, P., Kangas, A.J., Makinen, V.-P., Groop, P.-H., Savolainen, M.J., Juonala, M., Viikari, J.S., Kahonen, M., Lehtimäki, T., Raitakari, O.T. and Ala-Korpela, M., Characterization of systemic metabolic phenotypes associated with subclinical atherosclerosis, *Molecular BioSystems*, **2011**, 7(2): 385-393.
- Xia, J.G., Broadhurst, D.I., Wilson, M. and Wishart, D.S., Translational biomarker discovery in clinical metabolomics: an introductory tutorial, *Metabolomics*, **2013**, 9(2): 280-299.
- Xiao, C., Hao, F., Qin, X., Wang, Y. and Tang, H., An optimized buffer system for NMR-based urinary metabonomics with effective pH control, chemical shift consistency and dilution minimization, *Analyst*, **2009**, 134(5): 916-925.
- Xu, Q.-S. and Liang, Y.-Z., Monte Carlo cross validation, *Chemometrics and Intelligent Laboratory Systems*, **2001**, 56(1): 1-11.
-

- Xu, X.-H., Huang, Y., Wang, G. and Chen, S.-D., Metabolomics: a novel approach to identify potential diagnostic biomarkers and pathogenesis in Alzheimer's disease, *Neuroscience Bulletin*, **2012**, 28(5): 641-648.
- Zhang, X., Wang, Y., Hao, F., Zhou, X., Han, X., Tang, H. and Ji, L., Human Serum Metabonomic Analysis Reveals Progression Axes for Glucose Intolerance and Insulin Resistance Statuses, *Journal of Proteome Research*, **2009**, 8(11): 5188-5195.
- Zhao, X., Fritsche, J., Wang, J., Chen, J., Rittig, K., Schmitt-Kopplin, P., Fritsche, A., Häring, H.-U., Schleicher, E., Xu, G. and Lehmann, R., Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits, *Metabolomics*, **2010**, 6(3): 362-374.
- Zheng, P., Wang, Y., Chen, L., Yang, D., Meng, H., Zhou, D., Zhong, J., Lei, Y., Melgiri, N.D. and Xie, P., Identification and validation of urinary metabolite biomarkers for major depressive disorder, *Molecular & Cellular Proteomics*, **2013**, 12(1): 207-214.
- Zheng, X., Su, M., Pei, L., Zhang, T., Ma, X., Qiu, Y., Xia, H., Wang, F., Zheng, X., Gu, X., Song, X., Li, X., Qi, X., Chen, G., Bao, Y., Chen, T., Chi, Y., Zhao, A. and Jia, W., Metabolic signature of pregnant women with neural tube defects in offspring, *Journal of Proteome Research*, **2011**, 10(10): 4845-4854.
- Žitňanová, I., Korytár, P., Aruoma, O.I., Šustrová, M., Garaiová, I., Muchová, J., Kalnovičová, T., Püeschel, S. and Ďuračková, Z., Uric acid and allantoin levels in Down syndrome: antioxidant and oxidative stress mechanisms?, *Clinica Chimica Acta*, **2004**, 341(1-2): 139-146.
- Zuppi, C., Messana, I., Forni, F., Ferrari, F., Rossi, C. and Giardina, B., Influence of feeding on metabolite excretion evidenced by urine ¹H NMR spectral profiles: a comparison between subjects living in Rome and subjects living at arctic latitudes (Svaldbard), *Clinica Chimica Acta*, **1998**, 278(1): 75-79.
- Zuppi, C., Messana, I., Tapanainen, P., Knip, M., Vincenzoni, F., Giardina, B. and Nuutinen, M., Proton nuclear magnetic resonance spectral profiles of urine from children and adolescents with type 1 diabetes, *Clinical Chemistry*, **2002**, 48(4): 660-662.

Annex I: Main characteristics of prenatal and newborn health metabolomic studies

Table A-I.1- Summary of publications reported in literature of metabolomic in prenatal health research (including number of samples, gestational age at sampling, analytical technique used, statistical approach and main biochemical findings). Amino acids in three letter-code, 3-HBA: 3-hydroxybutyrate, BALF: bronchoalveolar lavage fluid, BSLR: backward stepwise logistic regression, CMM: Correlation matrix map, CS: cesarean section, CV: cross-validation, CVA: canonical variates analysis, GABA: γ -aminobutyrate, HCA: hierarchical cluster analysis, HIE: hypoxic ischemic encephalopathy, HIVA: hidroxyisovalerate, HMLR: hierarchical multiple linear regression, HPLC: high performance liquid chromatography, LC-HRMS: liquid chromatography high-resolution mass spectrometry, LDA: linear discriminant analysis, MCADD: medium chain acyl-coenzyme A dehydrogenase deficiency, NMND: *N*-methylnicotinamide, nS-HR-MS: nanospray high resolution MS, NTD: neural tube defect, OSC: orthogonal signal correction, GP: genetic programing, PE: preeclampsia, PKU: phenylketonuria, RF: random forest, ROC: receiver operating characteristic, TCA: tricarboxylic acid, TMA: trimethylamine, TMAO: trimethylamine-*N*-oxide, UPCA: unsupervised unfold principal component analysis, VD: vaginal delivery.

Biofluid Subject groups, <i>n</i>	Gestational age at sampling	Analytical technique	Statistical tools	Main biochemical findings	Reference
<i>Preeclampsia</i>					
Amniotic fluid PE <i>n</i> =7, Control <i>n</i> =36	3 rd T	NMR (600 MHz)	Univariate, LDA	higher acetate, succinate, choline and possibly citrate	Bock 1994
Placental extracts PE <i>n</i> =3, Control <i>n</i> =5	At delivery	MS, TLC-MS	—	Changes in phospholipid profiles	Jain 2004
Blood plasma PE <i>n</i> =87, Control <i>n</i> =87	At diagnosis	GC-MS	Genetic Programming, External validation	No biochemical interpretation	Kenny <i>et al.</i> 2005
Blood plasma PE <i>n</i> =11, Control, <i>n</i> =11	At diagnosis	NMR (500 MHz)	Univariate, PCA	Lower ketone bodies	Turner <i>et al.</i> 2007
Blood plasma PE <i>n</i> =11, Control <i>n</i> =11	At diagnosis	NMR (500 MHz)	Univariate, PCA	Altered aromatic amino acids	Turner <i>et al.</i> 2008
Blood plasma PE <i>n</i> =20, Control <i>n</i> =20	PE 186-272 (days) Control 186-282 (days)	UPLC-MS	Univariate, External validation, ROC	Ischemic injury and limited oxidative capacity	Kenny <i>et al.</i> 2008
Placental villous explants Control, <i>n</i> =11	At delivery	GC-MS	Univariate, PCA	Hypoxia related changes 2-deoxyribose, threitol erythritol and hexadecanoic acid	Heazell <i>et al.</i> 2008
Blood plasma PE <i>n</i> =15, Control <i>n</i> =22	PE: 36±3 g.w. Control: 39±1 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA	Higher concentrations of ala, gly and ergothioneine in plasma erythrocytes	Turner <i>et al.</i> 2009
Placental villous explants PE <i>n</i> =6, Control <i>n</i> =6	At delivery (g.w. not specified)	UPLC-MS	Univariate	Hypoxia related changes, namely gln and glu, trp metabolism and leukotriene or prostaglandin metabolism	Dunn <i>et al.</i> 2009

Blood plasma Test set: pre-PE $n=60$, Control $n=60$, Validation set: pre-PE $n=39$, Control $n=40$	2 nd T: 15±1 g.w.	UPLC-MS	Univariate, PLS-DA, External Validation, Permutation test, ROC, Genetic algorithm	14-metabolite signature comprising amino acids, carbohydrates, carnitines, dicarboxylic acids, fatty acids, ketones, keto- and hydroxyl- fatty acids, lipids, phospholipids and steroids	Kenny <i>et al.</i> 2010
Blood Serum pre-PE $n=41$, Control, $n=41$	1 st and 2 nd T: 11-14 g.w.	LC-MS/MS	Univariate, Logistic regression, ROC	Higher hydroxyhexanoylcarnitine, ala, phe and glu	Odibo <i>et al.</i> 2011
Placental extracts PE $n=6$, Control, $n=11$	At delivery: PE: 37 ⁺⁶ -41 ⁺⁰ g.w. Control: 35 ⁺² -39 ⁺³ g.w.	GC-MS UPLC-MS	Univariate, PCA	Altered mitochondrial metabolism, specifically vitamin D, oxidative and nitrate stress	Dunn <i>et al.</i> 2012
Blood serum Late onset pre-PE $n=30$, Early onset pre-PE $n=30$, Control $n=59$	1 st T: 11 ⁺⁶ -13 ⁺⁶ g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, Internal validation (permutation test, ROC)	Early-onset: changes in citrate, glycerol, HIBA and met, Late-onset: val, pyruvate, 3-HBA, 1-Me-histidine, glycerol and TMA	Bahado-Singh <i>et al.</i> 2012
Maternal urine Pre-PE $n=9$: Control $n=84$	2 nd T: 15-26 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, Variable Selection, Internal validation (MCCV), STOCYS	TCA cycle intermediates and precursors	Diaz <i>et al.</i> 2013
Fetal Malformations					
Amniotic fluid Open spina bifida $n=9$, Control $n=10$	2 nd T (g.w. not specified)	NMR (600 MHz)	Univariate, LDA	Lactate, glu, acetate	Bock 1994
Fetal urine Obstructive uropathy $n=14$	1 st to 3 rd T 13-30 g.w.	NMR (500 MHz)	—	Glycosuria, amino and organic aciduria	Foxall <i>et al.</i> 1995
Amniotic fluid Spina bifida $n=14$, Control $n=18$	Spina bifida 28.8±7 g.w., Control 24.7±12.2 g.w.	NMR (500 MHz)	Univariate, Multiple linear regression	Derangement in amino acid metabolism (higher succinate and glutamine)	Groenen <i>et al.</i> 2004
Amniotic fluid FM $n=12$, Control $n=51$	2 nd T 15-24 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (CV), STOCYS	Changes in glycolysis and gluconeogenesis, kidney underdevelopment (glucose, amino and organic acids)	Graça <i>et al.</i> 2009
Amniotic fluid FM $n=27$, Control $n=82$	2 nd T 15-22 g.w. (Control), 14-25 g.w. (disorders)	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, <i>i</i> PLS-DA, Internal validation (MCCV)	Enhanced glycolysis under hypoxia and reduced use of the respiratory chain, folic acid pool	Graça <i>et al.</i> 2010
Maternal urine FM $n=26$, Control $n=25$	2 nd T FM 14-25 g.w., Control 16-21 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (MCCV), STOCYS	enhanced gluconeogenesis under hypoxia, TCA cycle demand, derangements in methionine and nucleotide metabolism	Diaz and Pinto <i>et al.</i> 2011

Maternal blood plasma FM $n=27$, Control $n=20$	2 nd T FM 14-25 g.w. (+ 3 3 rd T), Control 16-21 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (MCCV), STOC SY	Energy metabolism, lipids conversion and methionine metabolism (acetone, 3-HBA, lipids, betaine)	Diaz and Pinto <i>et al.</i> 2011
Maternal blood serum NTD $n=101$, Control $n=143$	1 st T, 2 nd T and 3 rd T	UPLC-MS GC-MS	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (CV)	Impaired mitochondrial respiration, changes in the neurotransmitter GABA and changes in methionine metabolism	Zheng <i>et al.</i> 2011
Maternal blood serum NTD $n=80$, Control $n=95$	2 nd and 3 rd T >28 g.w. NTD $n=68$, Control $n=81$; ≤28 g.w.	UPLC-MS	Univariate, PCA, OPLS-DA, Internal validation (CV)	methionine cycle and folic acid pool, consistent with previous, perturbation in neurogenesis and oxidative stress	Liang <i>et al.</i> 2012
Amniotic fluid FM $n=22$, Control $n=26$	2 nd T FM 15-25 g.w., Control 16-22 g.w.	UPLC-MS	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (MCCV), SHY	Enhanced glycolysis under hypoxia, enhanced TCA and fetal kidney underdevelopment, lipids b-oxidation	Graça <i>et al.</i> 2012
Maternal urine FM $n=13$, Control $n=21$	2 nd T FM 15-25 g.w., Control 16-22 g.w.	UPLC-MS	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (MCCV), SHY	Gut microflora, glucuronic acid and sulphate conjugation reactions	Graça <i>et al.</i> 2012
Maternal urine FM $n=35$, Control $n=84$	2 nd T FM 14-24 g.w., Control 15-26 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, Internal validation (MCCV), STOC SY	Enhanced gluconeogenesis, lipid oxidation and ketone body synthesis, respiratory chain and nucleotides metabolism	Diaz <i>et al.</i> 2013b
Amniotic fluid FM $n=34$, Control $n=40$	2 nd T FM 14-25 g.w. Control 16-22 g.w.	MIR	Univariate, PCA, PLS-DA, Internal validation (MCCV), STOC SY, SHY	Decreased glucose (due to fetal hypoxia and disturbance of energy metabolism), changes in amino acids: gln, ile, leu, increased proteins	Graça <i>et al.</i> 2013
Preterm Delivery					
Newborn urine Preterm $n=5$, Control $n=6$	Shortly after birth Preterm 25-36 g.w., Control 38 g.w.	NMR (500 MHz)	—	Altered renal function (taurine, TMAO)	Foxall <i>et al.</i> 1995
Newborn urine Preterm $n=8$, Control $n=27$	Day 1 to Day 6 Preterm 28.9 ± 1.2 g.w., Control 39.4 ± 1.0 g.w.	NMR (600 MHz)	Univariate, K-nearest neighbour (kNN)	Tryptophan/NAD ⁺ pathway (taurine, NMND)	Trump <i>et al.</i> 2006
Amniotic fluid Test set: PTD without IAI $n=19$ PTD with IAI $n=20$, Control	2 nd and 3 rd T Test set 22-33 g.w. Validation set: 23-33	LC-MS	Univariate, LDA, External validation (RF, confusion matrices)	Changes in amino acids and sugar contents	Romero <i>et al.</i> 2010
Amniotic fluid Pre-PTD $n=12$, Control $n=82$	2 nd T 15-22 g.w. (Control), 14-25 g.w. (disorders)	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, <i>i</i> PLS-DA, Internal validation (MCCV)	Increased oxidative stress, amino acids and lung surfactants	Graça <i>et al.</i> 2010

Maternal urine pre-PTD <i>n</i> =17, Control <i>n</i> =25	2 nd T pre-PTD 16-21 g.w., Control 16-21 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (MCCV), STOC SY	Choline and amino acids metabolism	Diaz and Pinto <i>et al.</i> 2011
Cervicovaginal fluid pre-PTD (with short cervix) <i>n</i> =5, Control (term with short cervix) <i>n</i> =5	3 rd T pre-PTD with short cervix 31-33 g.w.,	UPLC-MS	PLS-DA	No biochemical interpretation	Auris-Blais <i>et al.</i> 2011
Newborn urine Preterm <i>n</i> =41, Control <i>n</i> =26	First 12 h of life Preterm 32.0±3 g.w., Control 39.7±0.8 g.w.	NMR (400 MHz)	Univariate, PCA, PLS-DA, HCA	Tyr, arg and proline metabolism, urea cycle, tyr, trp and phe biosynthesis	Atzori <i>et al.</i> 2011
Amniotic fluid pre-PTD <i>n</i> =11, Control <i>n</i> =26	2 nd T pre-PTD 16-21 g.w., Control 16-22 g.w.	UPLC-MS	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (MCCV), SHY	Placental dysfunction, changes in amino acids and sugars	Graça <i>et al.</i> 2012
Breast milk PTD <i>n</i> =20, Control <i>n</i> =3	1-13 weeks after delivery. Delivery at 26- 36 g.w.	NMR (500 MHz) GC-MS	PCA, Internal validation (CV)	Changes in carbohydrate over the first weeks of lactation	Marincola <i>et al.</i> 2012
Umbilical cord blood plasma and maternal blood plasma VLBW <i>n</i> =8, FT <i>n</i> =8	At delivery VLBW < 32 g.w., FT > 37g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, Internal validation (permutation tests)	Altered maternal-fetal-placental symbiosis. MB: acetate, lipids, pyruvate, gln, val and thr UCB: lipoproteins, pyruvate and	Tea <i>et al.</i> 2012
Umbilical cord blood plasma and maternal blood plasma VLBW <i>n</i> =7, FT <i>n</i> =8	At delivery VLBW < 32 g.w., FT > 37g.w.	LC-HRMS	Univariate, PCA, PLS-DA, Internal validation (permutation tests), HCA, UPCA	Changes in lipids, fatty acid oxidation, proteins metabolism, anti-oxidant defence, polyamine and puridine flux	Alexandre- Gouabau <i>et al.</i> 2013
Maternal urine Pre-PTD <i>n</i> =26, Control <i>n</i> =84	2 nd T pre-PTD 16-21 g.w., Control 15-26 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, Internal validation (MCCV), STOC SY	Enhanced lipid oxidation and oxidative stress	Diaz <i>et al.</i> 2013
Amniotic fluid Pre-PTD <i>n</i> =14, Control <i>n</i> =40	2 nd T Pre-PTD 16-21 g.w. Control 16-22 g.w.	MIR	Univariate, PCA, PLS-DA, Internal validation (MCCV), STOC SY, SHY	Decreased carboxylate-containing compounds, higher protein content, changes in citrate, ala, his, ile, leu, phe, tyr and val	Graça <i>et al.</i> 2013
Gestational Diabetes Mellitus and Maternal Hyperglycemia					
Amniotic fluid GDM <i>n</i> =5; Control <i>n</i> not specified	3 rd T (g.w. not specified)	NMR (600 MHz)	Univariate, LDA	No differences found	Bock 1994
Amniotic fluid Pre-diag. GDM <i>n</i> =27; Control <i>n</i> =82	2 nd T: Disorders: 14-25 g.w. Controls: 15-22 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, <i>i</i> PLS-DA, Internal validation (MCCV)	Hyperglycemia, amino acids and lipids metabolism, nucleotide biosynthesis and renal function	Graça <i>et al.</i> 2010

Maternal urine Pre-diag. GDM $n=29$; Control $n=25$	2 nd T: GDM: 16-22 g.w. Controls: 16-21 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA; Internal validation (MCCV), STOC SY	Nucleotide and amino acids metabolism, altered biotin status	Diaz and Pinto <i>et al.</i> 2011
Blood plasma Pre-diag. GDM $n=14$; Control $n=20$	2 nd T: GDM: 16-22 g.w. Controls: 16-21 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA; Internal validation (MCCV), STOC SY	Methionine metabolism	Diaz and Pinto <i>et al.</i> 2011
Amniotic fluid Pre-diag. GDM $n=23$; Control $n=26$	2 nd T: GDM: 16-21 g.w. Controls: 16-22 g.w.	UPLC-MS	Univariate, PCA, PLS-DA, OPLS-DA; Internal validation (MCCV), SHY	Decreased gly and increased sugar possibly glucose	Graça <i>et al.</i> 2012
Maternal urine Pre-diag. GDM $n=20$; Control $n=21$	2 nd T: GDM: 16-21 g.w. Controls: 16-22 g.w.	UPLC-MS	Univariate, PCA, PLS-DA, OPLS-DA; Internal validation (MCCV), SHY	Increased choline	Graça <i>et al.</i> 2012
Maternal urine $n=823$ participants	1 st -2 nd T 8-20 g.w. 2 nd T 28±2 g.w. 10-16 w. post partum	NMR (600 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (permutation tests)	Increased citrate possibly due to hyperglycemia	Sache <i>et al.</i> 2012
Maternal blood plasma Hyperglycemia $n=67$, Control $n=50$	2 nd and 3 rd T 24-36 g.w.	GC-MS	Univariate, Linear and logistic regression, bootstrapping, false discovery rate	Higher triglycerides, ketone bodies and amino acids, and low 1,5-anhydroglucitol	Scholtens <i>et al.</i> 2013
Maternal urine Pre-diag. GDM $n=42$, Control $n=84$	2 nd T: GDM: 16-22 g.w. Controls: 15-26 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, Variable Selection; internal validation (MCCV), STOC SY	Hyperglycemia, changes in amino acids and nucleotide metabolism, gut microflora	Diaz <i>et al.</i> 2013
Umbilical cord blood GDM $n=30$, Controls $n=40$	At delivery (> 37 g.w.)	NMR (600 MHz)	Univariate, PCA, OPLS-DA internal validation (MCCV)	Fetal hyperinsulinemia, increased oxidative stress and fetal glycolysis, changes in amino acids metabolism	Dani <i>et al.</i> 2013
Chromosomal disorders (CD)					
Amniotic fluid T21, $n=22$ Control $n=41$	2 nd T T21 15.7 g.w., Control 16.7 g.w.	GC-MS	Univariate	Purines, pyrimidines, creatinine, <i>N</i> -acetyl-aspartate, methylmalonic acid, malonic acid and other amino compounds	Baggot <i>et al.</i> 2008
Amniotic fluid T21 $n=24$, Controls $n=1257$	2 nd T 15-20g.w.	HPLC	Univariate, cluster analysis	Decreased amino acids, fetal kidney underdevelopment	Amorini <i>et al.</i> 2011
Maternal urine CD $n=23$, Control $n=25$	2 nd T CD 14-25 g.w., Control 16-21 g.w.		Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (MCCV), STOC SY	Changes due to hypoxia or lipids metabolism (choline)	Diaz and Pinto <i>et al.</i> 2011
Maternal blood plasma CD $n=23$, Control $n=20$	2 nd T CD 14-25 g.w., Control 16-21 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (MCCV), STOC SY	Changes in lipids profiles	Diaz and Pinto <i>et al.</i> 2011

Maternal blood serum T21 $n=30$, Control $n=60$	1 st T 11 ⁺⁰ -13 ⁺⁶ g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, Logistic regression analysis, Internal validation (Permutation tests, ROC, GP)	Increased oxidative stress and perturbation in brain growth and myelination	Bahado-Singh 2013a
Maternal blood serum T18 $n=30$, T21 $n=30$, Control $n=114$	1 st T 11 ⁺⁰ -13 ⁺⁶ g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, Logistic regression analysis, Internal validation (Permutation tests, ROC, GP)	Increased homocysteine transsulfuration, increased oxidative stress in T18. Higher TMA in T18 compared to T21	Bahado-Singh <i>et al.</i> 2013b
Maternal urine CD, $n=33$, T21, $n=13$, Control, $n=84$	2 nd T 15-26 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, Internal validation (MCCV), STOCSY	Increased ketone bodies and NMND excretion, TCA operation, amino acids, glucose, pyruvate. Perturbation in methionine metabolism in T21	Diaz <i>et al.</i> 2013
Perturbations of fetal growth					
Placental extracts SGA $n=9$, Control $n=8$	At delivery	UPLC-MS	Univariate, PCA, Internal validation (CVA, false discovery rate)	altered amino acid transfer into fetal circulation and fetoplacental perfusion	Horgan <i>et al.</i> 2010
Maternal blood plasma SGA $n=8$, Control $n=6$; 2 nd T pre-SGA $n=40$, Control $n=40$	At delivery 2 nd T 15 \pm 1 g.w.	UPLC-MS	Univariate, PLS-DA, Internal validation (ROC, permutation test, Genetic algorithm)	changes in lipids, carnitines and fatty acids due to placental dysfunction	Horgan <i>et al.</i> 2011
Newborn urine Preterm with IUGR $n=26$ Preterm with AGA $n=30$	Within 24 h of birth and at 96 h of life g.w. not specified	NMR (500 MHz)	Univariate, PCA, PLS-DA	perturbations in lipids and proteins metabolism, altered insulin secretion	Dessi <i>et al.</i> 2011
Umbilical cord and maternal blood plasma LBW $n=20$, Control $n=30$	At delivery GA>37 g.w.	NMR (600 MHz)	Univariate, PCA, PLS-DA, Internal validation (CV)	Changes in carbohydrate storage, impaired placental transfer of amino acids, changes in DNA methylation. No differences in	Ivorra <i>et al.</i> 2012
Umbilical vein serum IUGR, $n=22$, Control $n=21$	At delivery IUGR 32-41.3 g.w., Control 36.4-41.2 g.w.	LC-HRMS	Univariate, PCA, Cluster analysis, Internal validation (ROC)	22-metabolite signature suggestive of altered amino acid transfer, with emphasis to changes in phe, trp and glu	Favretto <i>et al.</i> 2012
Maternal urine pre-IUGR $n=10$, Control $n=84$	2 nd T pre-IUGR 16-24 g.w., Control 15-26 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, Internal validation (MCCV), STOCSY	Decreased amino acids	Diaz <i>et al.</i> 2013
Asphyxia, Respiratory Depression Syndrome and Meconium Aspiration Syndrome					
Newborn urine Asphyxia $n=20$, Healthy $n=10$	Not specified	NMR (400 MHz)	—	Increased lactate, changes in TCA cycle intermediates	Ma <i>et al.</i> 1995


Newborn urine Resuscitation and poor outcome $n=11$, Resuscitation and good outcome $n=13$, Control $n=216$	Resuscitation and poor outcome 260 ± 34 (days) Resuscitation and good outcome 276 ± 19 (days) Control 277 ± 10 (days) Collected 12 and 12-48h after birth	GC-MS	Univariate, HCA, Internal validation (ROC)	Increased organic acids	Chu <i>et al.</i> 2005
Newborn BALF RDS $n=4$	Birth before 34 g.w.	NMR (500 MHz), GC-MS	—	Changes in undecane, decanoic acid, dodecanoic acid, hexadecanoic acid, octadecanoic acid, hexadecanoic acid	Fabiano <i>et al.</i> 2011
Umbilical cord serum HIE $n=31$, Asphyxia without HIE $n=40$, Control $n=71$	HIE 40.3 ± 1.1 g.w., Control 1 40.3 ± 0.9 g.w., Asphyxia without HIE	Direct flow injection and LC-MS/MS	Univariate, PLS-DA, CVA, BSLR, Internal validation (permutation tests, ROC),	Changes in amino acids, acylcarnitines, and glycerophospholipids	Walsh <i>et al.</i> 2012
Umbilical cord serum HIE $n=25$, Asphyxia $n=34$, Controls $n=59$	Birth after 36 g.w.	NMR (600 MHz)	Univariate, CVA, internal validation (ROC)	Altered ketones linked to neurological injuries and hypoxia related changes	Reinke <i>et al.</i> 2013
Premature rupture of the membranes (PROM)					
Amniotic fluid pre-PROM $n=34$, Control $n=82$	2 nd T 15-22 g.w. (Control), 14-25 g.w. (disorders)	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, iPLS-DA, Internal validation (MCCV)	Subtle changes in met, gln and thr	Graça <i>et al.</i> 2010
Maternal urine pre-PROM $n=38$, Control $n=25$	2 nd T pre-PROM 15-21 g.w., Control 16-21 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (MCCV), STOCYS	No changes found	Diaz and Pinto <i>et al.</i> 2011
Maternal blood plasma pre-PROM $n=18$, Control $n=25$	2 nd T pre-PROM 15-21 g.w., Control 16-21 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, OPLS-DA, Internal validation (MCCV), STOCYS	Faint changes in acetate, gln, citrate and albumin	Diaz and Pinto <i>et al.</i> 2011
Maternal urine pre-PROM $n=68$, Control $n=84$	2 nd T pre-PROM 15-21 g.w., Control 15-26 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, Internal validation (MCCV), STOCYS	No changes found	Diaz <i>et al.</i> 2013
Amniotic fluid pre-PROM $n=37$, Control $n=40$	2 nd T Pre-PROM 16-23 g.w. Control 16-22 g.w.	MIR	Univariate, PCA, PLS-DA, Internal validation (MCCV), STOCYS, SHY	No biochemical interpretation	Graça <i>et al.</i> 2013
Inborn Errors of Metabolism					
Dried blood spots PKU $n=5$, Control $n=49$	PKU 1 st weeks Control not specified	NMR (400 MHz)	PCA, PLS-DA	Increased phe	Constantinou <i>et al.</i> 2004
Newborn urine PKU $n=10$, Control $n=47$	PKU: day 10 Control: at birth	NMR (400 MHz)	PCA, PLS-DA	Increased phe	Constantinou <i>et al.</i> 2005

Dried blood spots PKU $n=21$, MCADD $n=21$, other IEM $n=24$, Control $n=500$	Not specified	nS-HR-MS	PCA, LDA	No biochemical interpretation	Denes <i>et al.</i> 2012
Infection					
Newborn urine Bacterial infections $n=17$, Control $n=27$	Collected day1 to day 6 Bacterial infections 40.2 \pm 0.9 g.w., Control	NMR (600 MHz)	Univariate, K-nearest neighbour (kNN)	No changes found	Trump <i>et al.</i> 2006
Newborn urine CMV infected $n=12$, not- infected $n=11$	First 3 weeks of life	NMR (500 MHz)	PLS-DA, internal validation (CV, permutation test)	Altered energy metabolism and other virus related changes	Fanos <i>et al.</i> 2013
Other relevant studies					
Umbilical cord blood CS (without labor) $n=6$, VD (with medication) $n=16$, VD	CS 37.0-40.0 g.w., VD (with medication) 39.6- 41.4 g.w., VD (without	GC-MS	Univariate	Labor related stress and hormonal actions inducing changes in amino acids, sugars and ketone bodies	Hashimoto <i>et al.</i> 2012
Maternal blood serum Poor pregnancy outcome $n=40$, Control $n=40$	3 rd T Poor pregnancy outcome 28-41 g.w., Control 28-42 g.w.	UPLC-MS	Univariate, PCA	changes in lipids metabolism due to placental dysfunction	Heazell <i>et al.</i> 2012
Amniotic fluid Lung maturity: Immature $n=13$ Transitional $n=8$, Mature $n=22$,	2 nd and 3 rd T	NMR (600 MHz)	Univariate, LDA	Changes in gly, choline, creatinine, succinate, glu and ala	Bock 1994
Healthy Pregnancy Studies					
Amniotic fluid Healthy pregnancy, $n=92$	<15 g.w. $n=21$, 15-18 g.w. $n=24$, 19-23 g.w. $n=21$, 24-28 g.w. $n=15$,	GC-MS	—	Changes in succinic acid and gln	Ottolenghi <i>et al.</i> 2010
Amniotic fluid Healthy pregnancy, $n=78$	18 g.w. $n=18$, 19 g.w. $n=22$, 20 g.w. $n=17$, 21 g.w. $n=10$, 22 g.w. $n=11$	HPLC	Univariate, correlation analysis, HMLR	Changes in ala, gln, val, tau, leu, glu, tyr, phe, ile, and met concentration	Athanasiadis <i>et al.</i> 2011
Blood serum $n=54$ (highest FM prevalence) $n=40$ (average FM prevalence)	2 nd T High prevalence 9.5-23.6 g.w., average prevalence 14.4-18.4 g.w.	Immunoassays GC-MS	Univariate, PCA, PLS-DA, OPLS-DA, internal validation (CV, HCA, CMM)	Lower carbohydrates, lipids, higher amino acids and urea-cycle metabolites. Changes in trace elements (folate, vitamin D, Se, Zn, Cu, Sr, Cd, Pb)	Su <i>et al.</i> 2011
Placental extracts 1 st T, ≤ 8 g.w. $n=6$; ≥ 10 g.w. $n=6$ At delivery, VD $n=6$, CS $n=5$	1 st T: ≤ 8 g.w.: ≥ 10 g.w. At delivery: VD: 39 ⁺⁶ - 40 ⁺³ , CS: 37 ⁺⁶ -41 ⁺⁰	GC-MS UPLC-MS	Univariate, PCA	Lipids composition, suggesting changes in fatty acids β -oxidation for mitochondrial ATP. Higher oxidative stress in VD	Dunn <i>et al.</i> 2012


Maternal urine Healthy pregnant and NP 1 st T <i>n</i> =16, 2 nd T <i>n</i> =20, 3 rd T	1 st T: 11-13 g.w., 2 nd T: 15-26 g.w., 3 rd T: 29-39 g.w.	NMR (500 MHz)	Univariate, PCA, PLS-DA, internal validation (MCCV), STOC SY	21-metabolite signature comprising changes in amino acids, lipids oxidation, ketone bodies, methionine metabolism and	Diaz <i>et al.</i> 2013
Maternal urine Healthy pregnancies, <i>n</i> =83	1 st T 11 g.w.	NMR (600 MHz)	Univariate, PCA, OSC-PLS-DA	Higher oxidative stress and disturbed energy metabolism in higher pesticide exposure	Bonvallot <i>et al.</i> 2013

Annex II: Informed consents used in this study


Figure A-II.1: Informed consents given to pregnant women donating urine at the time of amniocentesis.




Centro Hospitalar de Coimbra
Maternidade Bissaya Barreto



universidade de aveiro



30 anos a projectar futuros



Universidade de Coimbra

Coloque aqui a
Etiqueta de identificação da consulta

**PEDIDO DE AUTORIZAÇÃO PARA RECOLHA DE
SANGUE, ÚRINA E EXCEDENTE DO LÍQUIDO AMNIÓTICO**

1. As amostras recolhidas destinam-se a análise confidencial no âmbito de um projecto de investigação sobre: “DIAGNÓSTICO PRECOCE DE DESORDENS NA SAÚDE DA GRÁVIDA E DO FETO POR MÉTODOS ESPECTROSCÓPICOS”. Esta investigação não trará quaisquer consequências para a mãe ou bebé podendo, no entanto, levar a importantes avanços na área de detecção precoce de doenças do feto e da mãe.

2. As amostras ficarão sob a responsabilidade da equipa da Universidade de Aveiro e seus colaboradores e a sua análise será feita em total confidencialidade. Os dadores deverão abdicar de quaisquer direitos a potenciais benefícios resultantes da investigação em curso podendo, a qualquer momento, negar o uso das amostras sem que seja necessária qualquer justificação e desde que a identificação das amostras seja ainda possível.

Autorizo a análise da minha urina: Sim ☐ Não ☐

Autorizo a análise do meu sangue: Sim ☐ Não ☐

Autorizo a análise do excedente de líquido amniótico: Sim ☐ Não ☐

P-U _____/_____
P-S _____/_____
P-LA _____/_____
Motivo: _____

Assinatura:.....Data:

✂.....

Se respondeu afirmativamente acima, agradecemos que responda às seguintes questões:

1- N° de semanas de gravidez:.....

2- Peso actual..... Peso anterior à gravidez..... Altura..... Tensão arterial 1° trim.....

3- Medicação durante a sua gravidez? Sim ☐ Não ☐ Especifique

4- - Fez recolha da amostra em jejum? Sim ☐ Não ☐ Após pequeno almoço uniformizado? Sim ☐ Não ☐
Se não, qual o pequeno almoço?.....

5- Antecedentes: Pré-Eclampsia ☐ Diabetes Gestacional ☐ Hipertensão ☐ Outros ☐
Especifique.....

6- Teve algum problema específico na gestação actual? Sim ☐ Não ☐ Especifique

7- Parto na MBB? Sim ☐ Não ☐

Muito obrigada pelo seu tempo e sua colaboração.

A Equipa de Investigação

Pedido de recolha na altura da amniocentese


P-U _____/_____
P-S _____/_____
P-LA _____/_____
Motivo: _____

Versão Abril 2011

Figure A-II.2: Informed consents given to pregnant women donating samples at each pregnancy trimester.


Para senhoras ACOMPANHADAS
durante a gravidez

Centro Hospitalar de Coimbra



universidade de aveiro 30 anos a projectar futuros

Versão Jan11



Universidade de Coimbra

Coloque aqui a
Vinheta de identificação da consulta

PEDIDO DE AUTORIZAÇÃO PARA ANÁLISE DE:

- URINA DA MÃE E DO BEBÉ

- SANGUE DA MÃE

a) As amostras recolhidas destinam-se a análise confidencial para um projecto de investigação sobre: “DIAGNÓSTICO PRECOCE DE DESORDENS NA SAÚDE DA GRÁVIDA, DO FETO E DO RECÉM-NASCIDO POR MÉTODOS ESPECTROSCÓPICOS”. Esta investigação não trará quaisquer consequências para mãe ou bebé podendo, no entanto, levar a importantes avanços na detecção precoce de doenças da grávida e do bebé.

b) A urina e, quando possível, o sangue da mãe deverão ser recolhidos nas seguintes alturas:

1. Em cada consulta no Centro Hospitalar, durante a gravidez
2. Na altura da amniocentese, se aplicável,
3. Na altura do parto.

c) A urina do bebé deverá ser recolhida durante a sua estadia no Centro Hospitalar.

d) Informam-se os pais que serão contactados por correio ou email, durante os 2 primeiros anos de vida do bebé, para recolha de informação sobre a saúde do bebé e sua correlação com a análise de urina.

e) As amostras recolhidas ficarão sob a responsabilidade da equipa da Universidade de Aveiro e seus colaboradores e a sua análise será feita em total confidencialidade. Os dadores deverão abdicar de quaisquer direitos a potenciais benefícios resultantes da investigação em curso podendo, a qualquer momento, negar o uso das amostras sem que seja necessária qualquer justificação e desde que a identificação das amostras seja ainda possível.

Autorizo a análise da minha urina:

Sim ☐ Não ☐

Autorizo a análise do meu sangue:

Sim ☐ Não ☐

Autorizo a análise da urina do meu bebé:

Sim ☐ Não ☐

Tomei conhecimento das condições acima descritas, as quais aceito: Sim ☐ Não ☐

Assinatura da mãe:.....Data:

✂.....

3º Trimestre

Nº do Processo da mãe..... Data da recolha:.....

1- Nº de semanas de gravidez:..... Peso actual..... Altura..... Tensão arterial 3º trim.....

2- Medicação? Sim ☐ Não ☐..... Problemas durante gravidez? Sim ☐ Não ☐.....

3- Fez recolha da amostra em jejum? Sim ☐ Não ☐ Após pequeno almoço uniformizado? Sim ☐ Não ☐

se não, qual o pequeno almoço?.....

✂.....

2º Trimestre

Nº do Processo da mãe..... Data da recolha:.....

1- Nº de semanas de gravidez:..... Peso actual..... Altura..... Tensão arterial 2º trim.....

2- Medicação? Sim ☐ Não ☐..... Problemas durante gravidez? Sim ☐ Não ☐.....

3- Fez recolha da amostra em jejum? Sim ☐ Não ☐ Após pequeno almoço uniformizado? Sim ☐ Não ☐

se não, qual o pequeno almoço?.....

✂.....

1º Trimestre

Nº do Processo da mãe..... Data da recolha:.....

1- Nº de semanas de gravidez:..... Peso actual..... Altura..... Tensão arterial 1º trim.....


2- Medicação? Sim ☐ Não ☐..... Problemas durante gravidez? Sim ☐ Não ☐.....

3- Fez recolha da amostra em jejum? Sim ☐ Não ☐ Após pequeno almoço uniformizado? Sim ☐ Não ☐


se não, qual o pequeno almoço?.....

Figure A-II.3: Informed consents given to pregnant women donating urine at the time of delivery and of their babies.

Maternidade Bissaya Barreto



universidade de aveiro



Universidade de Coimbra

Coloque aqui a
Vinheta de identificação da consulta

PEDIDO DE AUTORIZAÇÃO PARA ANÁLISE DE:

- URINA DA MÃE E DO BEBÉ
- SANGUE DA MÃE E DO CORDÃO UMBILICAL

a) As amostras recolhidas destinam-se a análise confidencial para um projecto de investigação sobre: “DIAGNÓSTICO PRECOCE DE DESORDENS NA SAÚDE DA GRÁVIDA, DO FETO E DO RECÉM-NASCIDO POR MÉTODOS ESPECTROSCÓPICOS”. Esta investigação não trará quaisquer consequências para mãe ou bebé podendo, no entanto, levar a importantes avanços na detecção precoce de doenças da grávida e do bebé.

b) A urina e, quando possível, o sangue da mãe e sangue do cordão umbilical deverão ser recolhidos na altura do parto.

c) A urina do bebé deverá ser recolhida durante a estadia na Maternidade

d) Informam-se os pais que serão contactados por correio ou email, durante os 2 primeiros anos de vida do bebé, para recolha de informação sobre a saúde do bebé e sua correlação com a análise de urina.

e) As amostras recolhidas ficarão sob a responsabilidade da Universidade de Aveiro, Universidade de Coimbra, Maternidade Bissaya Barreto e seus colaboradores e a sua análise será feita em total confidencialidade. Os dadores deverão abdicar de quaisquer direitos a potenciais benefícios resultantes da investigação em curso podendo, a qualquer momento, negar o uso das amostras sem que seja necessária qualquer justificação e desde que a identificação das amostras seja ainda possível.

Autorizo a análise da minha urina:

Sim ☐ Não ☐

Autorizo a análise do meu sangue:

Sim ☐ Não ☐

Autorizo a análise do sangue do cordão umbilical:


Sim ☐ Não ☐

Autorizo a análise da urina do meu bebé:

Sim ☐ Não ☐

Tomei conhecimento das condições acima descritas, as quais aceito:

Sim ☐ Não ☐

Assinatura da mãe:..... **Data:**


Nº do Processo da mãe _____ **Nº do processo do bebé** _____

1- Nº de semanas de gravidez à data do parto:.....

2 – Peso antes da gravidez: **Peso actual:**..... **Altura:**..... **Tensão arterial 3º trim:**.....

3- Medicação? Sim ☐ Não ☐ Especifique

4- Problemas específicos surgidos durante gravidez? Sim ☐ Não ☐
 Especifique.....

Muito obrigada pelo seu tempo e sua colaboração.

A Equipa de Investigação

Annex III: Multivariate model parameters and Monte Carlo Cross Validation (MCCV) results

Table A-III.1: MVA models parameters and MCCV results obtained when considering the whole ^1H NMR spectra (left) and the 21 relevant integrals (right) found for NP and pregnant women. CR true and CR perm: classification rate for true and permuted models, Spec: specificity, Sens: sensitivity, Q^2_m : median of true models.

PLS-DA of whole ¹ H NMR spectra					MVA of 21 selected integrals					
NP vs. 1 st T	R ² _X	R ² _Y	Q ²	LV	R ² _X	R ² _Y	Q ²	LV		
	0.274	0.907	0.727	2	0.449	0.905	0.848	2		
	MCCV				MCCV					
	% CR true	% CR perm	% Spec	% Sens	Q ² _m	% CR true	% CR perm	% Spec	% Sens	Q ² _m
	88	47	90	90	0.724	98	48	96	98	0.88
1 st T vs. 2 nd T	R ² _X	R ² _Y	Q ²	LV	R ² _X	R ² _Y	Q ²	LV		
	0.224	0.894	0.483	2	0.375	0.622	0.281	2		
	MCCV				MCCV					
	% CR true	% CR perm	% Spec	% Sens	Q ² _m	% CR true	% CR perm	% Spec	% Sens	Q ² _m
	70	49	64	75	0.324	78	49	68	89	0.43
2 nd T vs. 3 rd T	R ² _X	R ² _Y	Q ²	LV	R ² _X	R ² _Y	Q ²	LV		
	0.240	0.841	0.580	2	0.400	0.857	0.745	2		
	MCCV				MCCV					
	% CR true	% CR perm	% Spec	% Sens	Q ² _m	% CR true	% CR perm	% Spec	% Sens	Q ² _m
	83	47	82	86	0.416	99	48	100	96	0.77
NP, 1 st T, 2 nd T, 3 rd T	R ² _X	R ² _Y	Q ²	LV	R ² _X	R ² _Y	Q ²	LV		
	0.309	0.544	0.291	3	0.553	0.564	0.465	3		

Figure A-III.1: True and permuted model Q^2 distributions (left) and ROC plots (right) given by MCCV of the PLS-DA models obtained using all spectral data points for a,b) NP vs. 1st T, c,d) 1st T vs. 2nd T and e,f) 2nd T vs. 3rd T. TPR: true positive rate, FPR: false positive rate.

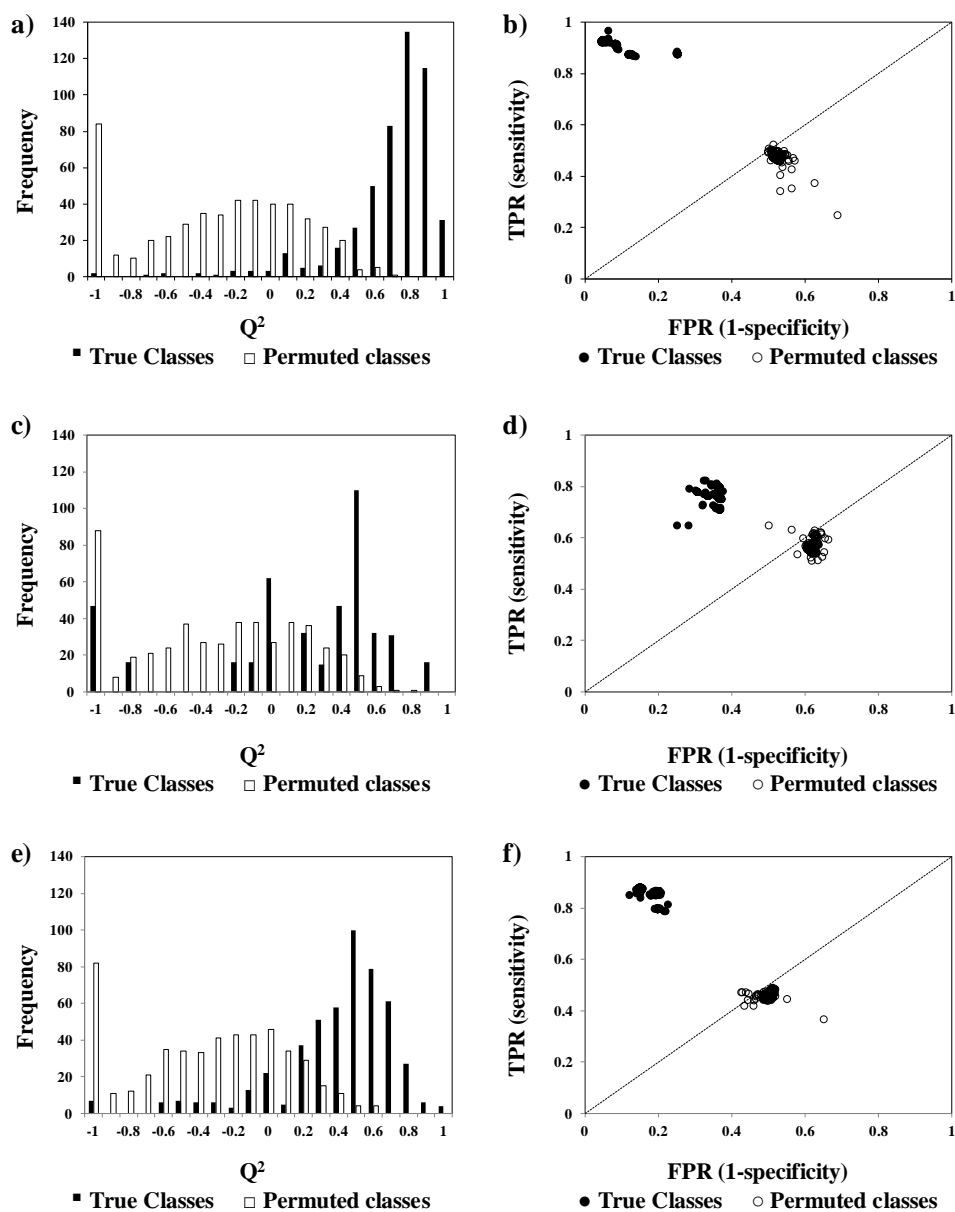


Figure A-III.2: True and permuted model Q^2 distributions (left) and ROC plots (right) given by MCCV of the PLS-DA models obtained using the selection of 21 relevant integrals for a,b) NP vs. 1st T, c,d) 1st T vs. 2nd T and e,f) 2nd T vs. 3rd T. TPR: true positive rate, FPR: false positive rate.

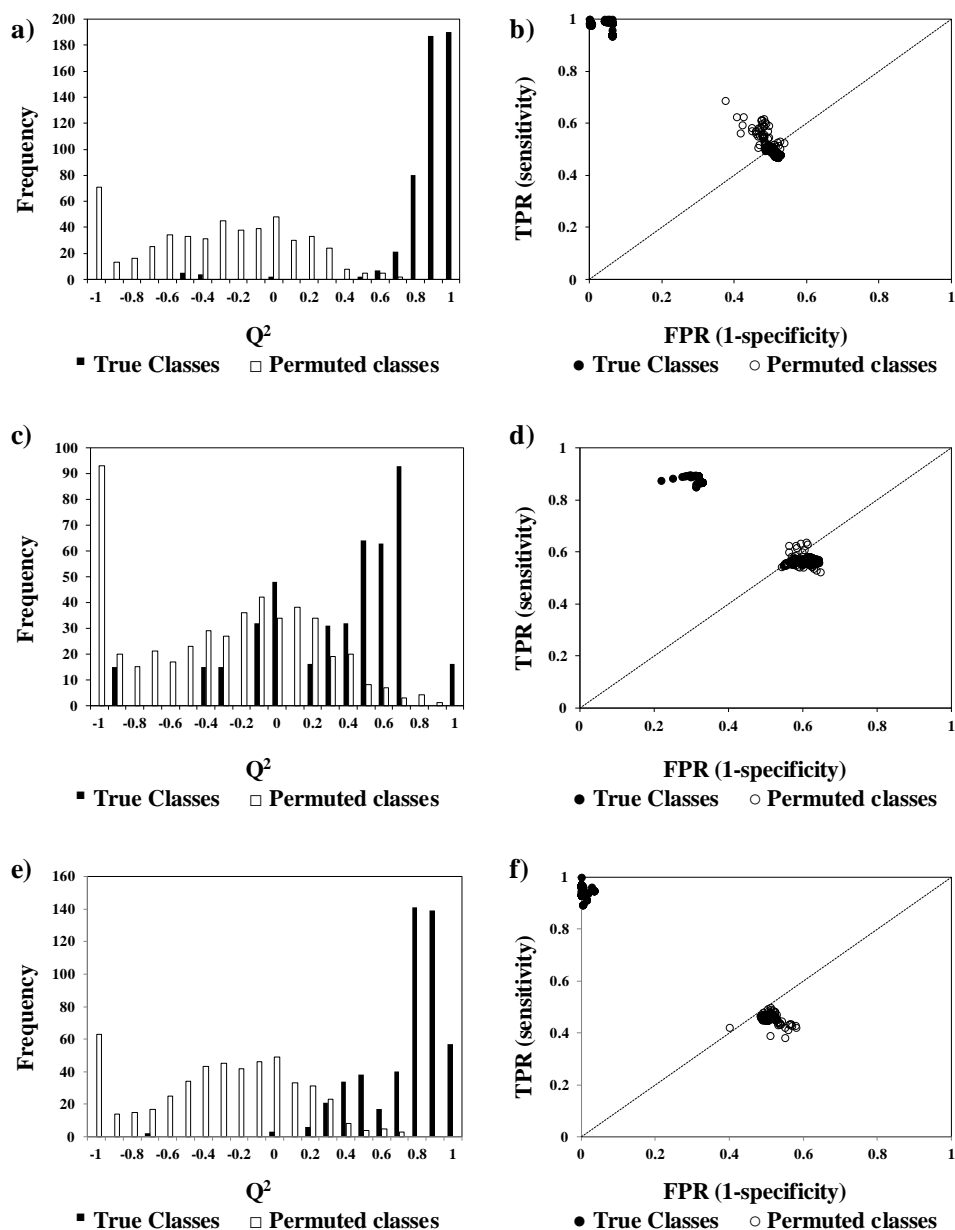


Table A-III.2: PLS-DA Q^2 (predictive power) and MCCV parameters for a) original dataset and for variables datasets selected by b) $VIP > 1$, c) $VIP/VIP_{cvSE} > 1$, d) $b/b_{cvSE} > 1$ and e) intersection of methods b, c and d. CR: classification rate, Sens: sensitivity, Spec: specificity, Q^2_m : median Q^2 obtained by MCCV. Best model results in bold.

Groups	Variable selection method	No. variables	CR (%)	Sens (%)	Spec (%)	Q^2	Q^2_{median}
FM	None	25591	69	49	75	0.251	0.065
	a) $VIP > 1$	10105	81	71	84	0.456	0.394
	b) $VIP/VIP_{cvSE} > 1$	12490	71	51	77	0.313	0.139
	c) $b/b_{cvSE} > 1$	11117	84	77	86	0.458	0.410
	d) Intersection of a, b and c	5743	84	74	87	0.444	0.368
CNS vs other FM	None	25591	62	69	44	0.186	0.647
	a) $VIP > 1$	9906	85	81	86	0.661	0.482
	b) $VIP/VIP_{cvSE} > 1$	20578	70	58	75	0.406	0.040
	c) $b/b_{cvSE} > 1$	10080	87	83	89	0.636	0.564
	d) Intersection of a, b and c	7176	87	83	90	0.705	0.647
CD	None	25591	68	34	73	-0.029	-0.112
	a) $VIP > 1$	10038	80	77	80	0.385	0.450
	b) $VIP/VIP_{cvSE} > 1$	20396	72	52	76	0.138	0.060
	c) $b/b_{cvSE} > 1$	8797	83	83	84	0.364	0.440
	d) Intersection of a, b and c	6132	85	87	84	0.428	0.530
T21 vs other CD	None	25591	61	66	51	0.146	-0.152
	a) $VIP > 1$	9765	81	77	83	0.665	0.578
	b) $VIP/VIP_{cvSE} > 1$	20589	68	63	71	0.395	0.017
	c) $b/b_{cvSE} > 1$	10481	91	90	91	0.645	0.569
	d) Intersection of a, b and c	7292	94	92	95	0.713	0.696
Pre-PTD	None	25591	73	32	77	0.014	-0.144
	a) $VIP > 1$	10390	80	67	82	0.348	0.308
	b) $VIP/VIP_{cvSE} > 1$	21215	75	41	87	0.167	-0.047
	c) $b/b_{cvSE} > 1$	8974	83	76	84	0.415	0.324
	d) Intersection of a, b and c	6323	84	79	84	0.458	0.422
Pre-PE	None	25591	91	23	93	0.170	-0.210
	a) $VIP > 1$	9374	93	65	93	0.412	0.287
	b) $VIP/VIP_{cvSE} > 1$	21363	92	33	93	-0.070	-0.070
	c) $b/b_{cvSE} > 1$	9521	93	55	93	0.466	0.342
	d) Intersection of a, b and c	5793	94	75	94	0.470	0.400
Pre-IUGR	None	25591	89	0	90	-0.013	-0.355
	a) $VIP > 1$	8945	90	81	90	0.360	0.343
	b) $VIP/VIP_{cvSE} > 1$	19366	88	1	89	0.180	-0.137
	c) $b/b_{cvSE} > 1$	8769	92	100	91	0.402	0.486
	d) Intersection of a, b and c	5627	94	99	94	0.438	0.540
Pre-PROM	None	25591	48	53	41	-0.145	-0.334
	b) $VIP > 1$	9638	63	60	66	0.232	0.215
	c) $VIP/VIP_{cvSE} > 1$	18989	54	48	57	0.033	-0.139
	d) $b/b_{cvSE} > 1$	8672	66	69	63	0.275	0.227
	e) Intersection of a, b and c	5810	70	68	72	0.344	0.341

Table A-III.3: PLS-DA Q^2 (predictive power) and MCCV parameters for a) original dataset and b) after variable selection for the GDM study. CR: classification rate, Sens: sensitivity, Spec: specificity, Q^2_m : median Q^2 obtained by MCCV.

Group	Dataset	No. variables	CR (%)	Sens (%)	Spec (%)	Q^2	Q^2_{median}
Diagnosed GDM vs. Control group 1	Original	25591	77	74	81	0.302	0.308
	Variable Selection	6968	96	93	100	0.750	0.703
<i>pre</i> -diagnosis GDM vs. Control group 2	Original	25591	66	43	72	0.061	0.069
	Variable Selection	6768	87	84	89	0.423	0.442
Treated GDM vs. Control group 3	Original	25591	65	64	65	0.141	0.069
	Variable Selection	6728	89	88	89	0.622	0.640
Insulin-requiring vs. Non-insulin-requiring	Original	25591	36	34	39	0.463	-0.263
	Variable Selection	7484	94	95	94	0.882	0.700
Insulin-treated vs. Non-insulin-treated	Original	25591	75	15	79	-0.027	-0.136
	Variable Selection	6243	94	100	93	0.674	0.702

Table A-III.4: PLS-DA Q^2 (predictive power) and MCCV parameters for a) original dataset and b) after variable selection. Mat: Maternal, CR: classification rate, Sens: sensitivity, Spec: specificity, Q^2_m : median Q^2 obtained by MCCV. Best model parameters are noted in bold.

	Group	Dataset	% original dataset	CR (%)	Sens (%)	Spec (%)	Q^2	Q^2_{median}
Newborn disorder	Preterm	Original	—	78	68	79	0.241	0.189
		Variable Selection	24.40%	92	92	92	0.556	0.603
	LGA	Original	—	71	47	75	0.220	0.009
		Variable Selection	28.86%	84	76	86	0.469	0.455
	PROM	Original	—	53	46	57	-0.025	-0.068
		Variable Selection	23.54%	77	75	79	0.406	0.377
	<i>pre</i> -Jaundice	Original	—	70	79	16	-0.089	-0.046
		Variable Selection	26.17%	78	45	84	0.117	0.254
	Respiratory depression	Original	—	84	71	85	0.147	0.052
		Variable Selection	27.88%	91	96	91	0.475	0.500
Maternal disorder	Malformations	Original	—	81	1	83	-0.006	-0.319
		Variable Selection	25.56%	89	73	90	0.424	0.391
	GDM	Original	—	71	15	76	-0.042	-0.203
		Variable Selection	23.82%	82	73	83	0.394	0.401
	Mat thyroid disease	Original	—	84	51	85	0.149	-0.130
		Variable Selection	22.60%	89	88	89	0.458	0.374
	Mat hypertension	Original	—	85	37	86	-0.193	-0.005
		Variable Selection	23.54%	90	100	90	0.432	0.485
	Mat psychiatric disease	Original	—	86	0	87	0.149	-0.068
		Variable Selection	26.90%	91	96	90	0.474	0.549
Confounding effects	Mat respiratory disease	Original	—	83	54	84	0.120	-0.024
		Variable Selection	28.07%	87	84	87	0.427	0.457
	Delivery mode	Original	—	64	52	71	0.039	0.130
		Variable Selection	25.92%	88	79	94	0.523	0.609
	Gender	Original	—	58	58	59	0.270	0.237
		Variable Selection	31.12%	76	75	76	0.474	0.515
	Day of life (Day 1 vs. day 2)	Original	—	76	80	7	0.069	-0.197
		Variable Selection	23.96%	86	87	73	0.452	0.466
	GA: 37 vs. 38 g.w	Original	—	59	69	38	0.126	0.060
		Variable Selection	31.08%	85	81	100	0.622	0.626
	GA: 38 vs. 39 g.w	Original	—	64	68	50	0.025	0.228
		Variable Selection	28.50%	87	89	82	0.494	0.650
	GA: 39 vs. 40 g.w	Original	—	51	39	57	0.031	-0.211
		Variable Selection	28.13%	78	74	82	0.526	0.525

Annex IV: Reproducibility of UPLC-MS experiments and model quality parameters

Figure A-IV.1: PCA score plots of UPLC-MS data set for a) HSS/ESI+, b) HSS/ESI-, c) HILIC/ESI+ and c) HILIC/ESI- using UV (left), Pareto (center) scaling and Log transformation (right). Study (\square) and QC (\bullet) samples.

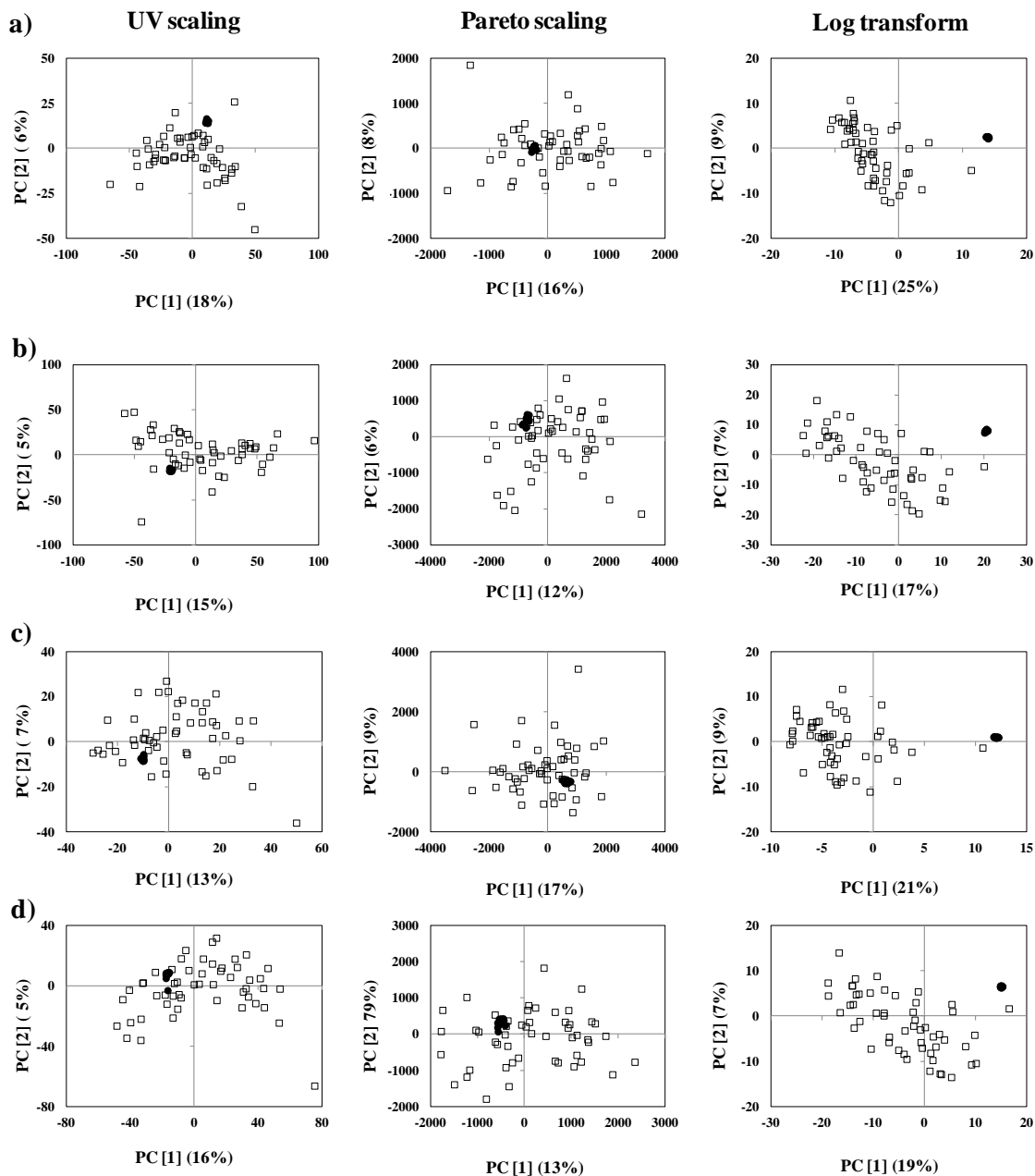


Table A-IV.1: PLS-DA Q^2 (predictive power) and MCCV parameters obtained for *pre*-PTD *vs.* controls analyzed by UPLC-MS with HSS/ESI+, HSS/ESI-, HILIC/ESI+ and HILIC/ESI- using Log transformation, Pareto and UV scaling using the intersection variable selection method. CR: classification rate, Sens: sensitivity, Spec: specificity.

Column/ Detection mode	Scaling/ Transform ation	Initial no. of features	No. features after variable selection	CR (%)	Sens (%)	Spec (%)	Q^2	Q^2_{median}
HSS/ESI+	Log		7870	97	96	97	0.730	0.738
	Pareto	42104	4413	95	92	99	0.602	0.753
	UV		12745	97	96	99	0.692	0.697
HSS/ESI-	Log		14039	96	96	96	0.643	0.662
	Pareto	78149	8474	59	62	59	-0.140	0.211
	UV		24515	98	97	99	0.666	0.718
HILIC/ESI+	Log		4156	86	85	87	0.562	0.603
	Pareto	22376	1894	91	92	90	0.417	0.651
	UV		6616	97	98	97	0.668	0.682
HILIC/ESI-	Log		8087	84	84	84	0.586	0.566
	Pareto	43223	2683	82	80	84	0.540	0.376
	UV		13171	93	91	95	0.607	0.628

Table A-IV.2: PLS-DA Q^2 (predictive power) and MCCV parameters obtained for *pre*-PTD *vs.* controls analyzed by UPLC-MS with HSS/ESI+, HSS/ESI-, HILIC/ESI+ and HILIC/ESI- using Log transformation, Pareto and UV scaling. Best model is shown in bold. CR: classification rate, Sens: sensitivity, Spec: specificity.

Column/ Detection mode	Scaling/ Transformation	No. features	CR (%)	Sens (%)	Spec (%)	Q^2	Q^2_{median}
HSS/ESI+	Log		69	69	70	0.258	0.242
	Pareto	2850	50	47	53	0.114	-0.051
	UV		54	57	52	0.298	-0.125
HSS/ESI-	Log		52	55	50	0.104	-0.084
	Pareto	7755	53	56	49	0.114	-0.179
	UV		53	56	51	0.328	-0.031
HILIC/ESI+	Log		65	65	64	0.004	0.121
	Pareto	1995	54	56	51	0.053	0.075
	UV		54	56	51	-0.016	0.039
HILIC/ESI-	Log		47	50	44	0.321	-0.104
	Pareto	4822	48	51	43	0.084	-0.203
	UV		51	54	48	0.121	-0.110